

# Spectral analysis and clustering of large stochastic networks. Application to the Lennard-Jones-75 cluster.

Maria Cameron<sup>1</sup> and Tingyue Gan<sup>1</sup>

December 8, 2015

## Abstract

We consider stochastic networks with pairwise transition rates of the form  $L_{ij} = a_{ij} \exp(-U_{ij}/T)$  where the temperature  $T$  is a small parameter. Such networks arise in physics and chemistry and serve as mathematically tractable models of complex systems. Typically, such networks contain large numbers of states and widely varying pairwise transition rates. We present a methodology for spectral analysis and clustering of such networks that takes advance of the small parameter  $T$  and consists of two steps: (1) computing zero-temperature asymptotics for eigenvalues and the collection of quasi-invariant sets, and (2) finite temperature continuation. Step (1) is reducible to a sequence of optimization problems on graphs. A novel single-sweep algorithm for solving them is introduced. Its mathematical justification is provided. This algorithm is valid for both time-reversible and time-irreversible networks. For time-reversible networks, a finite temperature continuation technique combining lumping and truncation with Rayleigh quotient iteration is developed. The proposed methodology is applied to the network representing the energy landscape of the Lennard-Jones-75 cluster containing 169,523 states and 226,377 edges. The transition process between its two major funnels is analyzed. The corresponding eigenvalue is shown to have a kink at the solid-solid phase transition temperature.

## 1 Introduction

The contemporary development of communications, information technologies and powerful computing resources has made networks a popular tool for data organization, representation and interpretation. In particular, networks have demonstrated their strong potential for modeling the dynamics of complex physical systems. These include time-reversible processes such as atomic or molecular cluster rearrangements and conformational changes in molecules [48, 43, 45], Markov State Models [35, 39, 16, 10, 29, 30, 34, 36], as well as time-irreversible processes such as walks of molecular motors [2]. Typically, such networks contain a large number of states or vertices ( $n = 10^p$ ,  $p = 4, 5, 6, \dots$ ), they are sparse and unstructured, and the pairwise transition rates vary by tens of orders of magnitude. As a result, analysis of the dynamics of such networks is a challenging problem due to their size, complexity, and severe floating point arithmetic issues.

Several approaches quantifying the dynamics of complex stochastic networks have been introduced. A. Bovier and collaborators developed the potential-theoretic approach and advanced

---

<sup>1</sup>University of Maryland, Department of Mathematics, College Park, MD 20742, [cameron@math.umd.edu](mailto:cameron@math.umd.edu), [tgan@umd.edu](mailto:tgan@umd.edu)

the mathematical spectral theory of metastability [4, 3, 5]. This theory is built upon an analogy with electric circuits. It is valid for time-reversible networks with an arbitrary form of pairwise transition rates. Its important result is sharp estimates for small eigenvalues and corresponding eigenvectors obtained under the assumption of the existence of spectral gaps.

The Transition Path Theory (TPT) originally proposed by W. E and E. Vanden-Eijnden [14] and further developed in [26, 9], like the potential-theoretic approach, was inspired by an analogy with electric circuits. However, there are important differences: (i) TPT does not assume time-reversibility and (ii) TPT is focused on the statistical analysis of so-called reactive trajectories. An alternative approach for analyzing reactive trajectories based on a set of recurrence relationships was proposed by M. Manhart and V. Morozov [27, 28].

Stochastic networks with pairwise transition rates of the form

$$L_{ij} = a_{ij} \exp(-U_{ij}/T), \quad (1)$$

where  $a_{ij} > 0$  and  $U_{ij} > 0$  are coefficients and  $T$  is a small parameter (typically, the absolute temperature in the physical context) arise as coarse-grained models of continuous systems justified by the Large Deviation Theory [18]. M. Freidlin proposed to describe the dynamics of such networks on a set of long time scales by means of a hierarchy of cycles [17, 19, 18] (we refer to them as Freidlin's cycles). A. Wentzell developed asymptotic estimates for eigenvalues [49]. In both cases, the results were given in the form of solutions of a series of optimization problems on a certain kind of directed graphs called W-graphs. Time-reversibility was not assumed. Freidlin's and Wentzell's approaches were further advanced by E. Olivieri and M. Vares [32].

The spectral analysis of stochastic networks with pairwise rates of the form of Eq. (1) was brought from the theoretical field to the computational field in [7, 8] for the case of time-reversible networks. The methodology presented in this work can be viewed as a two-fold extension of the one proposed in [7, 8]. First, a novel single-sweep algorithm for finding asymptotic estimates for eigenvalues via solving the series of the optimization problems on W-graphs is introduced. This algorithm does not require time-reversibility. Second, the finite-temperature continuation technique based on the Rayleigh quotient iteration [8] is empowered by a series of truncations and lumpings. This makes it applicable to networks with an arbitrary range of pairwise transition rates.

The dynamics of a stochastic network (also known as a continuous-time Markov chain) is determined by its generator matrix  $L$  and the initial probability distribution  $p_0$ . Throughout this paper, we assume that the spectral decomposition of  $L$  exists, i.e.,  $L = \Phi Z \Phi^{-1}$ , where  $Z$  is a diagonal matrix with eigenvalues of  $L$  along its diagonal, and columns of  $\Phi$  are the corresponding right eigenvectors. The spectral decomposition of the generator matrix gives a key to understanding the dynamics of the network, the extraction of quasi-invariant sets, and building coarse-grained models. However, its direct calculation for large and complex networks mentioned above might be exceedingly difficult due to issues related to floating-point arithmetic. Furthermore, even if the spectral decomposition is successfully computed, its interpretation might be not straightforward. Finally, for large networks, it is often desirable to extract only some particular eigenpairs associated with relaxation processes of interest rather than to obtain the whole set of eigenvalues and eigenvectors. Typically, the corresponding eigenvalues have relatively small, however, not necessarily the smallest, absolute values of their real parts. This means that, typically, the eigenvalues of physical interest are associated with some slowly but not necessarily the slowest decaying processes taking place in the system.

These issues can be reconciled by using the following two-step approach. Step one is the computation of zero-temperature asymptotic estimates for eigenvalues and a collection of quasi-invariant sets. The indicator functions of these quasi-invariant sets are asymptotic estimates for right eigenvectors in the time-reversible case [3]. We will show that each asymptotic eigenvalue is straightforward to interpret. Step two is the continuation of the eigenvalues describing relaxation processes of physical interest to a range of finite temperatures.

We apply the proposed approach to the stochastic network representing the energy landscape

of the Lennard-Jones cluster of 75 atoms<sup>1</sup>. For brevity, we will denote both, the Lennard-Jones cluster on  $N$  atoms and the network representing its energy landscape, by  $\text{LJ}_N$ . Vertices in this network correspond to local potential minima, while edges represent transition states between them. Transition rates between two adjacent minima are given by the Arrhenius law. Similar to the well-studied  $\text{LJ}_{38}$  cluster [42, 13, 31, 33, 6, 9, 7, 8], the energy landscape of  $\text{LJ}_{75}$  has a double-funnel structure (see Fig. 5 in Ref. [43] or Fig. 8.10(f) in Ref. [48]). The deep and narrow funnel is crowned with the global minimum which is the Marks decahedron with the point group  $D_{5h}$  [46]. This is minimum 1 in Wales's data set. There are several local minima based on icosahedral packing at the bottom of the wide and shallower funnel. One of them, minimum 92 in Wales's data set, is the second lowest one. The two deepest minima are shown in Fig. 1.

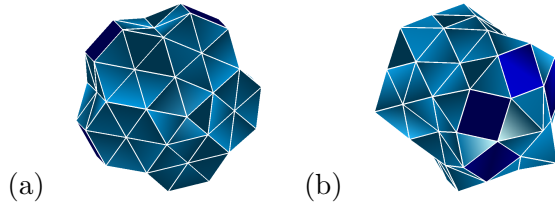


Figure 1: (a): The global potential minimum (minimum 1) of the  $\text{LJ}_{75}$  cluster, a Marks decahedron with the point group  $D_{5h}$ . (b): The second lowest potential energy minimum (minimum 92), a configuration based on the icosahedral packing with the point group  $C_1$  (no symmetry).

Besides similarities, there are considerable differences between  $\text{LJ}_{38}$  and  $\text{LJ}_{75}$ . First, the  $\text{LJ}_{75}$  network is significantly larger than  $\text{LJ}_{38}$  [44]. The  $\text{LJ}_{75}$  dataset contains 593,320 local minima and 452,315 transition states. The largest connected component of this network, containing the two deepest potential minima, 1 and 92, consists of 169,523 states (vertices) and 226377 undirected edges (excluding self-loops). The maximal vertex degree is 740. Second,  $\text{LJ}_{75}$  has a wider range of potential barriers that presents more severe numerical challenges than those encountered in the analysis of the  $\text{LJ}_{38}$  network [9, 8]. Third, unlike  $\text{LJ}_{38}$ ,  $\text{LJ}_{75}$  has extremely low solid-solid transition critical temperature. This leads to an interesting phenomenon that we advertise below and discuss in details in Section 7.2.4.

Thermodynamic properties of the  $\text{LJ}_{75}$  cluster were studied in [12, 25]. The process of physical interest in  $\text{LJ}_{75}$  is the transition process between the two main funnels, the Marks decahedron funnel and the icosahedral one. The solid-solid transition critical temperature of  $\text{LJ}_{75}$ , where the global potential energy minimum, minimum 1, gives place to icosahedral structures, is  $T_{ss}^{\text{LJ}_{75}} = 0.08$  reduced units [12]. It is very low in comparison with the potential energy barrier of 7.897 reduced units, separating minimum 1 from minimum 92. For comparison, the solid-solid transition critical temperature for  $\text{LJ}_{38}$  is  $T_{ss}^{\text{LJ}_{38}} = 0.12$  while the potential barrier separating the second lowest minimum from the global minimum is 3.543 [13, 24]. The melting temperature of the  $\text{LJ}_{75}$ , where another local maximum of the caloric curve is observed<sup>2</sup>, is  $T_{sl}^{\text{LJ}_{75}} = 0.25$ . Therefore, we will be interested in computing the eigenvalue and the eigenvector associated with the transition process between the two major funnels for the temperature range  $0 \leq T \leq 0.25$ . For brevity, we will denote this eigenvalue by  $\lambda(\text{ICO} - \text{MARKS})$ .

The eigenvalue  $\lambda(\text{ICO} - \text{MARKS})$  approximately equals the transition rate between the two major channels of the energy landscape of  $\text{LJ}_{75}$  [3]. We will show that the graph of  $\log(\lambda(\text{ICO} - \text{MARKS}))$  versus  $T^{-1}$  is nearly a piecewise-linear function with the kink at  $T_{ss}^{\text{LJ}_{75}} = 0.08$ . The slopes of the linear parts are in a good agreement with theoretical predictions. This means that the solid-solid transition critical temperature  $T_{ss}^{\text{LJ}_{75}} = 0.08$  lies within the range where the Large

<sup>1</sup>The data for the  $\text{LJ}_{75}$  network were kindly provided by Professor D. Wales, Cambridge University, UK.

<sup>2</sup>The value  $T_{sl}^{\text{LJ}_{75}} = 0.25$  corresponds to Wales's  $\text{LJ}_{75}$  dataset containing 593,320 local minima.

Deviation regime is valid with a peculiar twist: for  $T > T_{ss}^{LJ75} = 0.08$ , one needs to pretend that the icosahedral rather than the Marks decahedral funnel contains the global potential minimum.

The transition process between the two funnels is quantified by the corresponding eigencurrent. We will show that, despite the transition process becomes diverse as the temperature approaches the solid-liquid transition temperature  $T_{sl}^{LJ75} = 0.25$ , one can extract a sequence of edges along which the eigencurrent is highly concentrated. Remarkably, this sequence is not a subsequence of the asymptotic zero-temperature path (the MinMax path) as it is in the case of the LJ<sub>38</sub> cluster [8].

The rest of the paper is organized as follows. In Section 2, the significance of the spectral decomposition is discussed. In Section 3, the asymptotic estimates for eigenvalues are presented. In Section 4, nested properties of the optimal W-graphs are formulated. In Section 5, asymptotic estimates for left and right eigenvectors are discussed. The single-sweep algorithm for computing zero-temperature asymptotic estimates for eigenvalues and eigenvectors is introduced in Section 6. Section 7 is devoted to the application to LJ<sub>75</sub>. The upgraded finite-temperature continuation technique is explained in Section 7.2. Concluding remarks and Acknowledgements are in Sections 8 and 9 respectively. The proof of the nested properties of the optimal W-graphs is found in the Appendix.

## 2 Significance of the spectral decomposition

### 2.1 The general case

Suppose we have a stochastic network (a continuous-time Markov chain)  $(L, p_0)$ , where  $L$  is the generator matrix, and  $p_0$  is the initial probability distribution. The off-diagonal entries  $L_{ij}$  of  $L$  are the transition rates from states  $i$  to states  $j$ , while the diagonal entries  $L_{ii}$  are defined so that the row sums of  $L$  are zeros. The absolute values of the diagonal entries  $L_{ii}$  are the escape rates from states  $i$ . Throughout this work we assume that the Markov chain has a finite number of states  $n$  and is irreducible, i.e., there is a non-zero probability to reach any state from any other state. One can visualize the Markov chain using a directed weighted graph  $G(S, A, L)$ , where  $S$  is its set of states (vertices),  $|S| = n$ ,  $A$  is its set of arcs (directed edges), and  $L$  is the set of weights assigned to the arcs which is the set of all non-zero off-diagonal entries of the generator matrix  $L$  (we abuse notations here). Two vertices  $i$  and  $j$ ,  $i \neq j$ , are connected by an arc ( $i \rightarrow j$ ) if and only if  $L_{ij} \neq 0$  (i.e.,  $L_{ij} > 0$ ).

The time evolution of the probability distribution  $p(t) = [p_1(t), \dots, p_n(t)]$  is given by the master (or the Fokker-Planck) equation

$$\frac{dp}{dt} = pL, \quad p(0) = p^0. \quad (2)$$

Its solution can be readily written in terms of the spectral decomposition of  $L$ ,  $L = \Phi Z \Psi$ , where the columns of  $\Phi$  are the right eigenvectors, the rows of  $\Psi = \Phi^{-1}$  are the left eigenvectors, and  $Z$  is a diagonal matrix with the eigenvalues along its diagonal:

$$p(t) = p_0 \Phi e^{tZ} \Psi = \sum_{k=0}^{n-1} (p_0 \phi^k) e^{z_k t} \psi^k. \quad (3)$$

In Eq. (3), the solution  $p(t)$  is expanded in the basis of the left eigenvectors  $\psi_k$ . The coefficients of this expansion,  $(p_0 \phi^k) e^{z_k t}$ , are the projections of the initial distribution  $p_0$  onto the right eigenvectors  $\phi^k$  multiplied by the exponential functions  $e^{z_k t}$ . Given the above assumptions, it follows from the Perron-Frobenius theorem that there is a unique eigenvalue  $z_0 \equiv \lambda_0 = 0$  of  $L$ , and the rest of the eigenvalues  $z_k = -\lambda_k + i\mu_k$ ,  $k = 1, 2, \dots, n-1$ , have negative real parts. We order them so that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-1}.$$

The right eigenvector corresponding to  $\lambda_0 = 0$  is  $\mathbf{1} = [1, 1, \dots, 1]^T$  as the row sums of  $L$  are zeros, while left eigenvector corresponding to  $\lambda_0 = 0$  is the unique invariant probability distribution

$\pi = [\pi_1, \dots, \pi_n]$ . Using these notations and the fact that the distribution  $p_0$  sums up to 1, we rewrite Eq. (3) as

$$p(t) = \pi + \sum_{k=1}^{n-1} (p_0 \phi^k) e^{-\lambda_k t} e^{i\mu_k t} \psi^k. \quad (4)$$

Eq. (4) shows that for any initial probability distribution  $p_0$ , the probability distribution  $p(t)$  converges to the invariant distribution  $\pi$ , and the decay rate for  $k$ -th eigencurrent of  $p(t)$  is  $\lambda_k$ . Hence, on long time scales (say,  $t > t^*$ ), the probability distribution  $p(t)$  will be essentially determined only by those eigencurrents where  $\lambda_k$  is small (i.e.,  $\lambda_k < 1/t^*$ ). Since we are generally interested in the dynamics of the network on large times, it is important to be able to compute the eigenvalues with small real parts (in absolute values) and the corresponding left and right eigenvectors.

## 2.2 Time-reversible Markov chains

If the Markov chain is time-reversible, or, equivalently, its generator matrix  $L$  is in detailed balance with the invariant distribution  $\pi$ , i.e.,  $\pi_i L_{ij} = \pi_j L_{ji}$ , then the following additional properties of  $L$  hold.

- $L$  can be decomposed as

$$L = P^{-1}Q, \quad \text{where} \quad Q = Q^T \quad \text{and} \quad P = \begin{bmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_n \end{bmatrix}. \quad (5)$$

- $L$  is similar to a symmetric matrix

$$L_{sym} = P^{1/2} L P^{-1/2} = P^{-1/2} Q P^{-1/2}, \quad (6)$$

therefore, its eigenvalues are real and non-positive.

- The matrices  $\Psi$  and  $\Phi$  are related via

$$\Psi = \Phi^T P, \quad \text{i.e.,} \quad \psi^k = (\phi^k)^T P = [\pi_1 \phi_1^k, \dots, \pi_n \phi_n^k], \quad (7)$$

i.e., the right eigenvectors of  $L$  are orthonormal with respect to the inner product weighted by the invariant distribution, i.e.,  $\Phi^T P \Phi = I$ .

Relaxation processes in time-reversible Markov chains can be quantitatively described in terms of eigencurrents. A detailed discussion on it is found in [8]; also see [23, 41, 9]. Taking into account Eq. (7), the time evolution of the  $i$ -th component of the probability distribution can be rewritten as

$$\frac{dp_i}{dt} = (pL)_i = \sum_{j=1}^n p_j L_{ji} = \sum_{j \neq i} (p_j L_{ji} - p_i L_{ij}). \quad (8)$$

Plugging Eq. (7) in and using the detailed balance condition  $\pi_i L_{ij} = \pi_j L_{ji}$  we get

$$\frac{dp_i}{dt} = - \sum_{k=0}^{n-1} (p_0 \phi^k) e^{-\lambda_k t} \sum_{j \neq i} \pi_i L_{ij} (\phi_i^k - \phi_j^k) = - \sum_{k=0}^{n-1} (p_0 \phi^k) \sum_{j \neq i} F_{ij}^k, \quad (9)$$

where

$$F_{ij}^k(t) := e^{-\lambda_k t} \pi_i L_{ij} (\phi_i^k - \phi_j^k) \quad (10)$$

is the  $k$ -th eigencurrent along the arc  $(i \rightarrow j)$ . Note that  $c_k F_{ij}^k$  is the expectation of the difference of the numbers of transitions from  $i$  to  $j$  and from  $j$  to  $i$  performed by the system per unit time at time  $t$  if the initial distribution is

$$\pi + c_k \psi^k = (\pi_1(1 + c_k \phi_1^k), \dots, \pi_n(1 + c_k \phi_n^k)).$$

(The coefficient  $c_k$  is introduced to ensure that all entries of  $\pi + c_k \psi^k$  are nonnegative.)

Now we go over some properties of eigencurrents. Since  $\phi_i^0 = 1$ ,  $i \in S$ , the eigencurrent  $F^0$  associated with the zero eigenvalue  $z_0 = 0$  is zero everywhere. Furthermore,  $F_{ij}^k = -F_{ji}^k$ , as immediately follows from the definition. It is easy to verify that the sum of eigencurrent  $F^k$  at vertex  $i$  is

$$\sum_{j \neq i} F_{ij}^k = e^{-\lambda_k t} \lambda_k \pi_i \phi_i^k. \quad (11)$$

Therefore, the eigencurrent  $F^k$  is not, in general, conserved at vertex  $i$  but is either emitted, if  $\phi_i^k > 0$ , or absorbed, if  $\phi_i^k < 0$ . Hence, the set of states can be partitioned into emitting (more precisely, non-absorbing) and absorbing states:  $S = S_+^k \cup S_-^k$ , where

$$S_+^k := \{i \in S : \phi_i^k \geq 0\} \quad \text{and} \quad S_-^k = \{i \in S : \phi_i^k < 0\}.$$

This partition of the network is called *the  $k$ -th emission-absorption cut* or, briefly, *the  $k$ -th EA-cut*. The set of edges with endpoints in different components of this partition is also called the EA-cut.

It is shown in [8] that out of all cuts of the given network, the total eigencurrent  $F^k$  flowing through the edges of the  $k$ -th EA-cut is maximal. It is convenient to normalize the eigencurrent  $F^k$  so that its flux through the  $k$ -th EA-cut is unit or 100%. Then one can find the percentages of the eigencurrent  $F_{ij}^k$  for every edge  $(i, j)$  and hence obtain a quantitative description of the relaxation process starting from the initial distribution  $\pi + c_k \psi^k$ . In Section 7.2.5, we will analyze the eigencurrent corresponding to the transition process between the two major funnels of LJ<sub>75</sub>.

### 3 Asymptotic estimates for eigenvalues

A relationship between the zero-temperature asymptotics for eigenvalues of the generator matrix  $L$  with off-diagonal entries of the order of  $L_{ij} \asymp \exp(-U_{ij}/T)$  and a series of optimization problems on W-graphs was established by A. Wentzell in 1972 [49, 18]. A W-graph with  $k$  sinks for the given directed weighted graph  $G(S, A, U)$  is defined as a subgraph of  $G$  obtained as follows: (a) select a set of  $k$  vertices and call them sinks; (b) choose a subset of  $n - k$  arcs (directed edges) so that (i) there is a single outgoing arc from every non-sink vertex and (ii) the graph has no cycles. Note that a W-graph with  $k$  sinks contains  $n - k$  arcs. In particular, the W-graph with  $n$  sinks has no arcs. An optimal W-graph with  $k$  sinks, denoted by  $g_k^*$ , is the one for which the sum of weights of its arcs is minimal possible. I.e.,

$$g_k^* = \arg \min_{g \in G_k} \sum_{(i \rightarrow j) \in g} U_{ij}, \quad 1 \leq k \leq n, \quad (12)$$

where  $G_k$  is the set of all W-graphs on  $G(S, A, U)$  with  $k$  sinks. An example of a graph with four vertices and some of its W-graphs with two sinks are shown in Fig. 2.

Wentzell's asymptotic estimates for eigenvalues are up to the exponential order [49, 18]. For networks with pairwise transition rates of the form of Eq. (1), i.e.,  $L_{ij} = a_{ij} \exp(-U_{ij}/T)$ , Wentzell's result can be upgraded so that the estimates for the eigenvalues include not only exponents but also pre-exponential factors (pre-factors). Let us define the graph  $G(S, A, U)$  for the given generator matrix  $L$ , where the set of weights  $U = \{U_{ij}\}$  is the set of the exponential coefficients from Eq. (1). Throughout the rest of the paper we will adopt the following *genericness assumption*:

**Assumption 1.** All optimal W-graphs  $g_k^*$ ,  $1 \leq k \leq n$ , for the graph  $G(S, A, U)$  are unique.

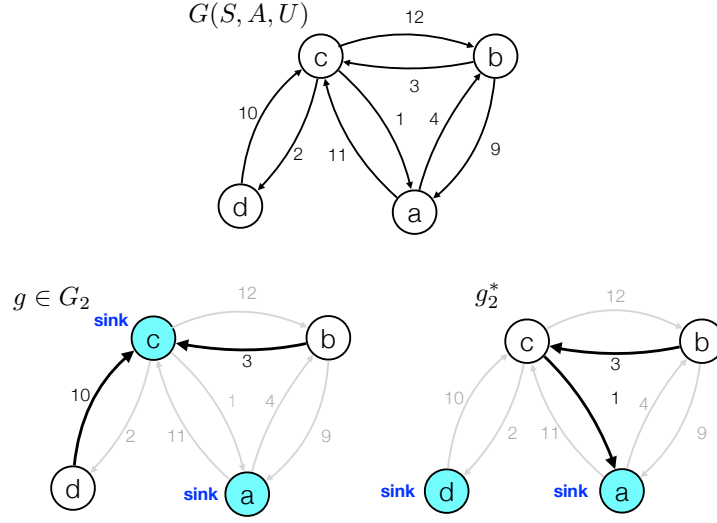


Figure 2: An example of a graph  $G(S, A, U)$  (top) and two W-graphs for  $G(S, A, U)$  with two sinks (bottom): a W-graph with two sinks (bottom left) and the optimal W-graph with two sinks (bottom right).

Then all eigenvalues of  $L$  are real and distinct and are approximates by

$$\lambda_k = A_k e^{-\Delta_k/\epsilon} (1 + o(1)), \quad \text{where} \quad (13)$$

$$\Delta_k = \sum_{(i \rightarrow j) \in g_k^*} U_{ij} - \sum_{(i \rightarrow j) \in g_{k+1}^*} U_{ij}, \quad (14)$$

$$A_k = \frac{\prod_{(i \rightarrow j) \in g_k^*} a_{ij}}{\prod_{(i \rightarrow j) \in g_{k+1}^*} a_{ij}}. \quad (15)$$

Below we provide a sketch of the proof of Eqs. (13)-(15).

Eqs. (13)-(15) are obtained from the consideration of the characteristic polynomial

$$\rho_L(z) = z^n + \beta_{n-1} z^{n-1} + \beta_{n-2} z^{n-2} + \dots + \beta_1 z = z(z - z_1) \dots (z - z_{n-1})$$

of the generator matrix  $L$ . It follows from our construction of the optimal W-graphs by Algorithm 1 presented in Section 6.2 below, that the uniqueness assumption of the optimal W-graphs implies that

$$0 < \Delta_{n-1} < \Delta_{n-2} < \dots < \Delta_1.$$

For this case, it is proven in [49] that all eigenvalues are real and distinct, i.e.,  $z_k = -\lambda_k$ , where  $\lambda_k > 0$ ,  $1 \leq k \leq n-1$ . One can show that the coefficients of the characteristic polynomial  $\rho_L(z)$  are given by

$$\beta_{n-k} = \sum_{1 \leq i_1 \neq \dots \neq i_k \leq n-1} \prod_{l=1}^k \lambda_{i_l} = \sum_{g \in G_k} \left( \prod_{(i \rightarrow j) \in g} a_{ij} e^{-U_{ij}/T} \right). \quad (16)$$

The summands in Eq. (16) range exponentially. Hence, each sum is dominated by its largest term. In the sum over the W-graphs with  $k$  sinks, the largest term is achieved on the optimal W-graph  $g_k^*$  which is unique by our assumption. The sum of products of all combinations of  $k$  numbers  $\lambda_{i_l}$ ,  $1 \leq l \leq k$ , where all  $1 \leq i_1, \dots, i_k \leq n-1$  are distinct, is dominated by

$\lambda_{n-1} \dots \lambda_{n-k}$ . Therefore,

$$\lambda_k = \frac{\lambda_{n-1} \dots \lambda_k}{\lambda_{n-1} \dots \lambda_{k+1}} \approx \frac{\prod_{(i \rightarrow j) \in g_k^*} a_{ij} e^{-U_{ij}/T}}{\prod_{(i \rightarrow j) \in g_{k+1}^*} a_{ij} e^{-U_{ij}/T}},$$

and hence Eqs. (13)-(15) hold.

For the time-reversible networks, the asymptotic estimates for the eigenvalues can be simplified. In this case, we assume that the off-diagonal entries of  $L$  are of the form

$$L_{ij} = \frac{b_{ij}}{b_i} e^{-(V_{ij}-V_i)/T}, \quad \text{where } V_{ij} = V_{ji} \text{ and } b_{ij} = b_{ji} \text{ for all } i, j \in S. \quad (17)$$

In other words, there exist the potential  $V$  and the pre-factor function  $b$  defined on all vertices and all edges such that  $U_{ij} = V_{ij} - V_i$  and  $a_{ij} = b_{ij}/b_i$ . Then the invariant probability distribution is readily calculated and given by

$$\pi_i = b_i e^{-V_i/T}, \quad 1 \leq i \leq n. \quad (18)$$

It was proven in [7] that for stochastic networks with pairwise rates of the form of Eq. (17), all edges (with erased directions) belonging to the optimal W-graph  $g_{k+1}^*$ , also belong to the optimal W-graph  $g_k^*$ , and all sinks of  $g_k^*$  are also sinks of  $g_{k+1}^*$ . Hence,  $g_k^*$  is obtained from  $g_{k+1}^*$  by removing one sink and adding one edge. We denote the disappearing sink and the newly added edge by  $s_k^*$  and  $(p_k^* \rightarrow q_k^*)$  respectively. Taking all this into account, it is easy to calculate that the exponent and the pre-factor for the asymptotic estimate of the  $k$ -th eigenvalue  $\lambda_k$ :

$$\Delta_k = V_{p_k^* q_k^*} - V_{s_k^*}, \quad A_k = \frac{b_{p_k^* q_k^*}}{b_{s_k^*}}. \quad (19)$$

## 4 Nested properties of optimal W-graphs

As we have discussed in Section 3, in order to obtain asymptotic estimates for eigenvalues of the generator matrix  $L$  of size  $n \times n$ , one needs to find the collection of optimal W-graphs  $g_k^*$ ,  $1 \leq k \leq n$ , i.e., to solve the collection of optimization problems on graphs given by Eq. (12). In order to find an optimal W-graph  $g_k^*$  with  $k$  sinks, one needs to minimize the sum in Eq. (12) with respect to (i) the choice of  $k$  sinks out of  $n$  vertices, and (ii) the choice of arcs so that each non-sink vertex has exactly one outgoing arc and these arcs create no cycles. Optimization problem (12) is hard to solve by brute force because the number of W-graphs for large number of vertices  $n$  is typically enormous. Fortunately, the collection of optimal W-graphs  $\{g_k^*\}_{k=1}^n$  under the assumption that all of them are unique, possesses nested properties allowing us to find them recursively starting from  $g_n^*$  and finishing with  $g_1^*$ .

**Theorem 1.** (Nested properties of optimal W-graphs) *Let  $G(S, A, U)$  be a weighted directed graph with  $n$  vertices such that all optimal W-graphs  $g_k^*$ ,  $1 \leq k \leq n$  are unique. Then for all  $1 \leq k < n$  the following nested properties hold.*

1. *Every sink of  $g_k^*$  is also a sink of  $g_{k+1}^*$ ;*
2. *Let  $S_k$  be the set of vertices in the connected component of  $g_{k+1}^*$  containing no sink of  $g_k^*$ . The collection of outgoing arcs from the subset of vertices  $S \setminus S_k$  in  $g_{k+1}^*$  coincides with the collection of outgoing arcs from the subset of vertices  $S \setminus S_k$  in  $g_k^*$ .*
3. *There is a single outgoing arc in  $g_k^*$  with tail in  $S_k$  and head in  $S \setminus S_k$ .*

**Remark** The collections of outgoing arcs from the subset of states  $S_k$  in  $g_{k+1}^*$  and  $g_k^*$  do not necessarily coincide.



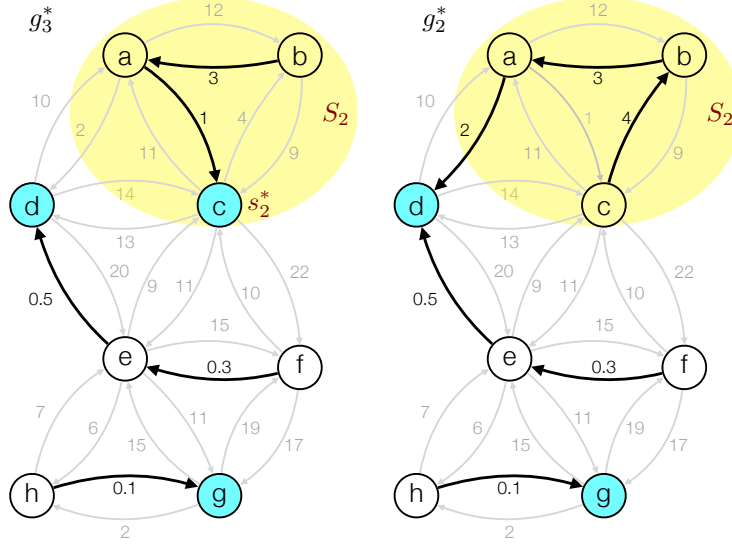


Figure 3: An illustration for the nested properties (Theorem 1).

A proof of Theorem 1 is found in the Appendix. The nested properties stated in Theorem 1 are illustrated in Fig. 3. The sets of arcs of the optimal W-graphs  $g_3^*$  (left) and  $g_2^*$  (right) with tails not in  $S_2 = \{a, b, c\}$  coincide. All sinks of  $g_2^*$  are also sinks of  $g_3^*$ . There is a single outgoing arc ( $a \rightarrow d$ ) in  $g_2^*$  with tail in  $S_2$  and head in  $S \setminus S_2 = \{d, e, f, g, h\}$ .

Theorem 1 can be compared to Theorem 3.5 in [7] stating the nested properties of the optimal W-graphs for the time-reversible case. In the time-reversible case, where  $U_{ij} = V_{ij} - V_i$ ,  $V_{ij} = V_{ji}$ ,  $i, j \in S$ , Claims 2 and 3 of Theorem 1 can be amplified as follows. Let  $\mathcal{T}_k^*$  and  $\mathcal{T}_{k+1}^*$  be the optimal forests obtained from the optimal W-graphs  $g_k^*$  and  $g_{k+1}^*$  by erasing the directions of their arcs. Then the forest  $\mathcal{T}_{k+1}^*$  is a sub-forest of  $\mathcal{T}_k^*$ . In addition, all optimal forests  $\mathcal{T}_k^*$ ,  $1 \leq k \leq n$ , are subgraphs of the minimum spanning tree

$$\mathcal{T}^* = \arg \min_{\mathcal{T}} \sum_{(i,j) \in \mathcal{T}} V_{ij},$$

(see Theorem 3.4 in [7]). Here, the minimum is taken over all spanning trees for the graph  $G(S, E, V)$ , where vertices  $i$  and  $j$  are connected by an edge  $(i, j)$  whenever there is an arc ( $i \rightarrow j$ ) or ( $j \rightarrow i$ ) in  $G(S, A, U)$ , and the weights  $V_{ij}$  are taken from  $U_{ij} = V_{ij} - V_i$ .

## 5 Asymptotic estimates for eigenvectors

In this paper, we limit our discussion on asymptotic estimates for right and left eigenvectors to the time-reversible case. Sharp asymptotic estimates for the right eigenvectors for time-reversible Markov chains were obtained in [3] (see also [4]). It follows from the theory developed in [3] that the  $k$ -th right eigenvector of  $L$  is approximated by the indicator vector  $\phi_0^k = [\phi_0^k(1), \dots, \phi_0^k(n)]^T$  of the quasi-invariant set  $S_k$  defined in Theorem 1, i.e.

$$\phi_0^0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \phi_0^k(i) = \begin{cases} 1, & i \in S_k \\ 0, & i \notin S_k \end{cases}, \quad 1 \leq k \leq n-1. \quad (20)$$

Asymptotic estimates for left eigenvectors of  $L$  can be deduced from those for the right ones. Since the matrix whose rows are left eigenvectors is the inverse of the matrix whose columns are

the corresponding right eigenvectors, we expect that the matrix of the asymptotic estimates for left eigenvectors is the inverse of the matrix

$$\Phi_0 := [\phi_0^0 \ \dots \ \phi_0^{n-1}].$$

Below we will define a matrix  $\Psi_0$  and show that it is the inverse of  $\Phi_0$ .

Let  $S_k$  be the set of vertices in the connected component of the optimal W-graph  $g_{k+1}^*$  containing no sink of  $g_k^*$ . Let  $s_k^*$  be the sink of  $S_k$ . According to Theorem 1, there exists a unique arc  $(p_k^*, q_k^*)$  in  $g_k^*$  such that  $p_k^* \in S_k$  and  $q_k^* \in (S \setminus S_k)$ . Let  $T_k$  be the connected component of  $g_{k+1}^*$  such that  $q_k^* \in T_k$ . Its sink will be denoted by  $t_k^*$  respectively. In the example in Fig. 3,  $k = 2$ ,  $S_2 = \{a, b, c\}$ ,  $s_2^* = c$ ,  $(p_2^*, q_2^*) = (a, d)$ ,  $T_2 = \{d, e, f\}$ , and  $t_2^* = d$ .

We introduce row vectors  $\psi_0^k = [\psi_0^k(1), \dots, \psi_0^k(n)]$  given by

$$\psi_0^k(i) = \begin{cases} 1, & i = s_k^* \\ -1, & i = t_k^* \\ 0, & \text{otherwise,} \end{cases}, \quad 1 \leq k \leq n-1, \quad \psi_0^0(i) = \begin{cases} 1, & i = s_0^* \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

Let  $\Psi_0$  be the matrix with rows  $\psi_0^k$ . We claim that  $\Psi_0$  is the inverse of  $\Phi_0$ , i.e.,

$$\Psi_0 \Phi_0 = \Phi_0 \Psi_0 = I. \quad (22)$$

Let us prove Eq. (22). We have

$$\psi_0^0 \phi_0^r = \phi_0^r(s_0^*) = \begin{cases} 1, & r = 0, \\ 0, & 1 \leq r \leq n-1, \end{cases} \quad (23)$$

and

$$\psi_0^k \phi_0^r = \phi_0^r(s_k^*) - \phi_0^r(t_k^*), \quad 1 \leq k \leq n-1, \quad 0 \leq r \leq n-1, \quad (24)$$

If  $k = r$ ,  $s_k^* \in S_k$  and  $t_k^* \notin S_k$ . Hence

$$\psi_0^k \phi_0^k = 1, \quad 1 \leq k \leq n-1.$$

If  $k < r$  then  $s_k^* \notin S_r$  and  $t_k^* \notin S_r$ . Hence for all  $i \in S$ ,  $\phi_0^r(s_k^*) = \phi_0^r(t_k^*) = 0$ . If  $k > r$  then the sinks  $s_k^*$  and  $t_k^*$  lie in the same connected component of  $g_{r+1}^*$ . Hence  $\phi_0^r(s_k^*) = \phi_0^r(t_k^*)$ . Therefore, for  $k \neq r$ , Eq. (24) implies that  $\psi_0^k \phi_0^r = 0$ . This completes the proof.

## 6 The single-sweep algorithm

In this Section, we introduce a single-sweep algorithm for computing the collection of optimal W-graphs, suitable for both time-reversible and time-irreversible networks. The presentation of this algorithm is given in the form convenient for programming.

### 6.1 Building the hierarchy of the optimal W-graphs

We will build the hierarchy of the optimal W-graphs  $g_k^*$  from bottom to top, i.e., starting from  $g_n^*$  containing  $n$  sinks (every vertex is a sink of  $g_n^*$ ) and no arcs, and finishing with  $g_1^*$  containing  $n-1$  arcs and one sink. Due to the nested properties (see Theorem 1) we know that we will need to remove one sink in a time, possibly rearrange arcs in the connected component containing the sink being removed, and connect it to some other connected component with a single arc. The algorithm with all its nuances will be written in the form of a pseudocode in Section 6.2. In this Section, we will explain some crucial parts of the algorithm such as the next arc selection procedure and the update rule (25) for weights  $U_{ij}$  and pre-factors  $a_{ij}$ .

In a W-graph, there can be at most one outgoing arc from each vertex. Therefore, in order to find the hierarchy of the optimal W-graphs, we start with picking the outgoing arcs of minimum

weight from each vertex. We proceed as follows. Let  $B_i$  be the set of outgoing arcs from the vertex  $i$ ,  $1 \leq i \leq n$ . In each set  $B_i$ , sort the arcs according to their weights  $U_{ij}$ , find the outgoing arc of minimal weight, and denote it by  $\text{min\_arc}(i)$ . Then define the set

$$M := \bigcup_{i \in S} \text{min\_arc}(i).$$

Now we start building optimal W-graphs. Find the arc  $(i \rightarrow j)$  of minimal weight in the set  $M$ , remove it from  $M$ , and define the set  $G^* = \{(i \rightarrow j)\}$  of arcs that has been removed from  $M$ . The arcs from  $G^*$  will be used for the construction of optimal W-graphs. Define the optimal W-graph  $g_{n-1}^*$  with one arc  $(i \rightarrow j)$  and  $n - 1$  sinks  $S \setminus \{i\}$ . Then find the next minimal weight arc  $(x \rightarrow y)$  in  $M$ , remove it from  $M$  and add to  $G^*$ . Suppose  $(x \rightarrow y) \neq (j \rightarrow i)$ , i.e., it does not create a cycle with the arc  $(i \rightarrow j)$ . Then define the optimal W-graph  $g_{n-2}^*$  with two arcs  $(i \rightarrow j)$  and  $(x \rightarrow y)$  and  $n - 2$  sinks  $S \setminus \{i, x\}$ . Suppose that, proceeding in this manner, the minimal weight arcs removed from  $M$  are such that they do not form any cycles. In this case, we will build the whole hierarchy of optimal W-graphs in  $n - 1$  steps, and the numbers  $\Delta_k$  and  $A_k$  defining the asymptotic estimates for the eigenvalues will be

$$\Delta_k = U_{p_k^*, q_k^*}, \quad A_k = a_{p_k^* q_k^*}, \quad n - 1 \geq k \geq 1.$$

However, such a scenario can occur only if we are particularly lucky with the network. Typically, sooner or later, the next minimal weight arc removed from the set  $M$  will create a directed cycle with some arcs in  $G^*$ . Hence, we cannot define the next optimal W-graph by adding this arc to the previous one. We need to develop another procedure to handle this case.

Suppose the optimal W-graphs  $g_n^*$ , ...,  $g_k^*$  are constructed as described above. Let the endpoints of the next minimal weight arc  $(x \rightarrow y)$  removed from  $M$  lie in the same connected component of  $g_k^*$ . Since, by construction, the set  $M$  contains at most one outgoing arc from each vertex, the new arc  $(x \rightarrow y)$  creates a simple directed cycle  $C$  with some other arcs in  $g_k^*$ :

$$C = \{x \rightarrow y \rightarrow \dots \rightarrow x\}.$$

Observe that, after adding arc  $(x \rightarrow y)$ , every vertex in the connected component  $S_C$  of  $g_k^*$  containing the cycle  $C$  has an outgoing arc. Furthermore, observe that  $x$  is a sink of  $g_k^*$ , as, prior to adding the arc  $(x \rightarrow y)$ , the vertex  $x$  have had no outgoing arc in  $g_k^*$ . In order to merge the component  $S_C$  with some other connected component, we need to add an arc  $(i \rightarrow j)$  such that  $i \in C$  and  $j \in (S \setminus C)$  to the set  $M$ . The fact that  $i$  must lie in  $C$  follows from Theorem 1. Indeed, suppose that the arc  $(i \rightarrow j)$  is removed from  $M$  after the optimal W-graph  $g_l^*$  for some  $l < k$  has been built, while  $g_{l-1}^*$  has not been built yet. First, for simplicity, we assume that no other cycle except for  $C$  has been created up to this point. Then, in order to build  $g_{l-1}^*$ , we will replace the arc  $\text{min\_arc}(i) \in C$  with the arc  $(x \rightarrow y) \in C$  and add the arc  $(i \rightarrow j)$  as shown in Fig. 4. Then the connected component of  $g_l^*$  containing  $i$  will remain connected in  $g_{l-1}^*$ , and every vertex will have at most one outgoing arc in  $g_l^*$ . The addition of the arc  $(i \rightarrow j)$  increases the sum of weights by  $U_{ij} + U_{xy} - U_{\text{min\_arc}(i)}$ .

Motivated by these considerations, we update the weights and pre-factors of the outgoing arcs with tails in the cycle  $C$  according to the following update rule: for all  $(i \rightarrow j) \in B_i$ , for all  $i \in C$  except for  $x$ , we set

$$U_{ij} = U_{ij} + U_{xy} - U_{\text{min\_arc}(i)}, \quad a_{ij} = a_{ij} a_{xy} / a_{\text{min\_arc}(i)}. \quad (25)$$

Note that we do not need to update the weights of the arcs emanating from  $x$ . Indeed, the last added arc with tail at  $x$  is  $x \rightarrow y \equiv \text{min\_arc}(x)$ . Hence the update rule (25) will not change the weights and pre-factors of the arcs emanating from  $x$ . Then we merge all sets  $B_i$  where  $i \in C$  into one set  $B$ . Next, we select the arc  $(p \rightarrow q) \in B$  of minimum weight (the effective weight), remove it from  $B$ , add it to  $M$ , and for each  $i \in C$  we set  $\text{min\_arc}(i) = (p \rightarrow q)$  and  $B_i = B$ . After that, we again remove the arc of minimum weight from  $M$  and add it to  $G^*$ .

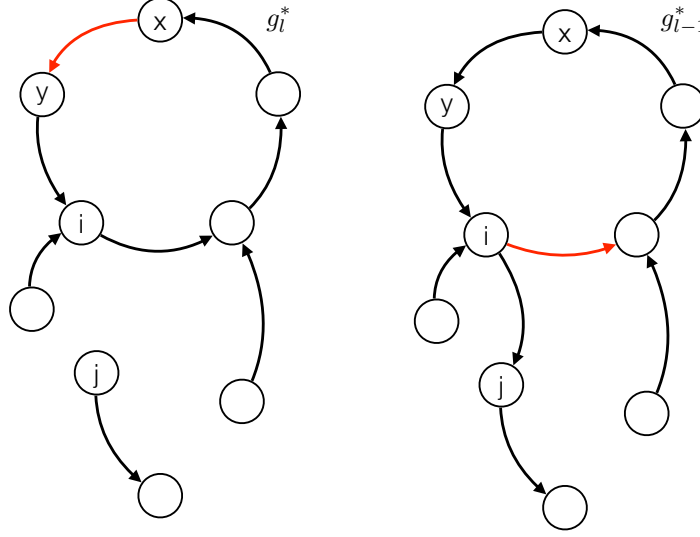


Figure 4: An illustration for the arc replacement in a cycle and the update rule (25). Black and red arcs belong to the set  $G^*$ . Black arcs form optimal W-graphs. The arc  $(x \rightarrow y)$  creates a cycle with other arcs in  $G^*$ .

Suppose that again, the next minimal weight arc  $(v \rightarrow w)$  removed from  $M$  creates a simple directed cycle  $C$  with some other arcs in  $G^*$ . We update the arc weights in all sets  $B_i$ ,  $i \in C$ , according to the update rule (25) and merge them to one set  $B$ . We call all previously created cycles containing vertices of  $C$  sub-cycles of  $C$ . Recursively, all sub-cycles of any sub-cycle of  $C$  will be also called sub-cycles of  $C$ . We denote the union of the cycle  $C$  and all its sub-cycles by  $C'$ . Note that, by induction, the set  $B$  is the union of all arcs emanating from all vertices of the cycle  $C'$ . Then we find the arc  $(p \rightarrow q)$  of minimal weight in  $B$ , remove it from  $B$ , add it to  $M$ , and set  $\text{min\_arc}(i) = (p \rightarrow q)$  and  $B_i = B$  for all  $i \in C$ .

Proceeding according to these recipes, we build the whole hierarchy of the optimal W-graphs  $g_k^*$  and obtain the collections of numbers  $\Delta_k$ ,  $A_k$ , collections of sinks  $s_k^*$  and  $t_k^*$ , collections of quasi-invariant sets  $S_k$  and cycles  $C_k$ . The cycle  $C_k$  is the cycle  $C$  (including all their sub-cycles) containing sink  $s_k^*$  that had been most recently created before the optimal W-graph  $g_k^*$  was built (and hence the vertex  $s_k^*$  lost its sink status). If the sink  $s_k^*$  has not been a part of any cycle created before  $g_k^*$  is built, then we set  $C_k = \{s_k^*\}$ . The collection of cycles  $C_k$  is a subset of the set of Freidlin's cycles [17, 19, 18]. Further, we, abusing notations, will denote the set of vertices in the cycle  $C$  also by  $C$ .

## 6.2 A pseudocode

The algorithm is summarized in the pseudocode below.

**Algorithm 1.** *A single-sweep algorithm constructing the hierarchy of the optimal W-graphs.*

**Input:** For each vertex  $i \in S$ , list the set of outgoing arcs with their weights  $U_{ij}$  and pre-factors  $a_{ij}$ ;

**Output:** The collections of numbers  $A_k$ ,  $\Delta_k$ , the collections of sets  $S_k$  and  $C_k$ , the disappearing sinks  $s_k^*$ , and the absorbing sinks  $t_k^*$ , the collection of exit arcs  $(p_k^* \rightarrow q_k^*)$ , and the collection of the optimal W-graphs  $g_k^*$  for  $n - 1 \leq k \leq 1$ ;

// **Initialization**

1. Denote the set of outgoing arcs from vertex  $i$  by  $B_i$ . Sort the arcs in each set  $B_i$  according

to their weights  $U_{ij}$  in the increasing order;

**2.** For each vertex  $i \in S$ , denote the outgoing arc of minimal weight by  $\text{min\_arc}(i)$ . Remove it from the set  $B_i$ . Create the set of minimum outgoing arcs  $M = \bigcup_{i \in S} \text{min\_arc}(i)$ . Set  $C(i) = \{i\}$ ;  $G^* = \emptyset$ ;

**3.** Set  $k = n - 1$ ;

// The main cycle

**while**  $\{|M| > 1\}$

**4.** Remove the minimum weight arc  $(x \rightarrow y)$  from the set  $M$  and add it to the set  $G^*$ ;

**if**  $\{x \text{ and } y \text{ belong to different connected components of } g_{k+1}^*\}$

**5.** Set  $(p_k^* \rightarrow q_k^*) = (x \rightarrow y)$ . The sinks  $s_k^*$  and  $t_k^*$  are the sinks of the connected components of  $g_{k+1}^*$  containing the vertices  $x$  and  $y$  respectively. The set  $S_k$  is the set of vertices in the connected component containing  $s_k^*$ . Define  $C_k = C(s_k^*)$ ;

**6.** Set  $\Delta_k = U_{xy}$  and  $A_k = a_{xy}$ ;

**7.** Find the connected component of the optimal W-graph  $g_k^*$  that results from merging of two connected components of  $g_{k+1}^*$  containing vertices  $x$  and  $y$  as follows. Starting from the sink of the connected component of  $g_{k+1}^*$  containing the vertex  $y$ , trace incoming arcs from the set  $G^*$  backwards until all vertices of this connected component are reached and every vertex except for the sink has a single outgoing arc;

**8.** Set  $k = k - 1$ ;

**else**

// a cycle is created;

**9.** Detect the cycle  $C$ ;

**10.** For each vertex  $i \in C$ , except for  $i = x$ , update the weights and the pre-factors of the outgoing arcs in the set  $B_i$  according to Eq. (25);

**11.** Merge sets  $B_i$  for all  $i \in C$ , i.e., define  $B := \bigcup_{i \in C} B_i$ . Set  $C' = \bigcup_{i \in C} C(i)$ ;

**while**  $\{\text{the vertices } r \text{ and } t, \text{ where } (r \rightarrow t) := \arg \min_{(p \rightarrow q) \in B} U_{pq}, \text{ belong to } C'\}$

**12.** Discard the arc  $(r \rightarrow t)$  from  $B$ ;

**end while**

**13.** Set  $B_i = B$  and  $\text{min\_arc}(i) = \arg \min_{(p \rightarrow q) \in B} U_{pq}$  for all  $i \in C'$ . Set  $C(i) = C'$  for all  $i \in C'$ ;

**14.** Delete the minimum outgoing arc from  $B$  and add it to  $M$ ;

**end if - else**

**end while**

Algorithm 1 was motivated by (i) Kruskal's algorithm [22, 1] for finding a minimum spanning tree in an undirected weighted graph, and (ii) Chu - Liu/Edmonds' algorithm [11, 15] designed for finding a spanning arborescence of a minimum weight with the root at the prescribed vertex (i.e., the directed analog of the minimum spanning tree).

### 6.3 An illustrative example

An example in Fig. 5 demonstrates how Algorithm 1 works. The given continuous-time Markov chain is shown in Fig. 5(a). For reader's convenience, we have labeled statements in Algorithm 1 with bold arabic numbers. We will refer to them as we work out the example. **1:** First, we sort the sets of outgoing arcs for each vertex and get:

$$B_a = \{(a \rightarrow b; 1), (a \rightarrow d; 1.8), (a \rightarrow e; 2), (a \rightarrow c; 5)\};$$

$$B_b = \{(b \rightarrow c; 3), (b \rightarrow a; 10)\};$$

$$B_c = \{(c \rightarrow a; 2), (c \rightarrow b; 2.6), (c \rightarrow d; 3.2)\};$$

$$B_d = \{(d \rightarrow a; 0.5), (d \rightarrow c; 1.5), (d \rightarrow e; 3)\};$$

$$B_e = \{(e \rightarrow d; 6), (e \rightarrow a; 7)\}.$$

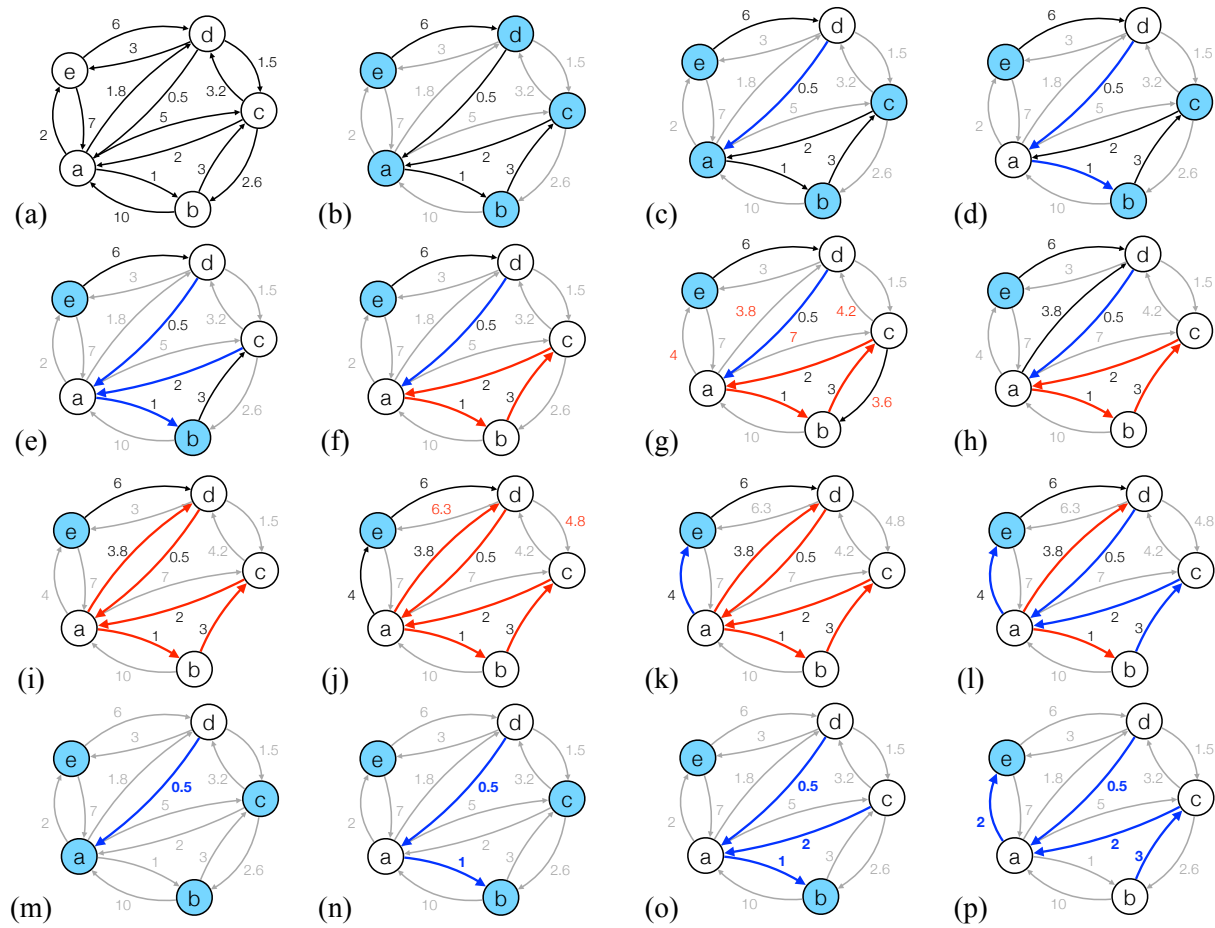


Figure 5: An example demonstrating how Algorithm 1 works.

**2:** Next, we remove the minimal weight arc from each set  $B_i$ ,  $i \in \{a, b, c, d, e\}$ , form the set  $M$  out of them and sort the arcs in  $M$ :

$$\begin{aligned} B_a &= \{(a \rightarrow d; 1.8), (a \rightarrow e; 2), (a \rightarrow c; 5)\}; \\ B_b &= \{(b \rightarrow a; 10)\}; \\ B_c &= \{(c \rightarrow b; 2.6), (c \rightarrow d; 3.2)\}; \\ B_d &= \{(d \rightarrow c; 1.5), (d \rightarrow e; 3)\}; \\ B_e &= \{(e \rightarrow a; 7)\}; \\ M &= \{(d \rightarrow a; 0.5), (a \rightarrow b; 1), (c \rightarrow a; 2), (b \rightarrow c; 3), (e \rightarrow d; 6)\}; \\ C(a) &= \{a\}, C(b) = \{b\}, C(c) = \{c\}, C(d) = \{d\}, C(e) = \{e\}; \end{aligned}$$

The arcs constituting the set  $M$  are the thick black ones in Fig. 5(b). **3:** Set  $k = 4$ .

Now the main cycle starts. **4:** The minimal weight arc  $(d \rightarrow a; 0.5)$  is removed from  $M$  and added to  $G^*$ . Since  $d$  and  $a$  belong to different connected components of  $g_5^*$  (containing all 5 sinks and no arcs), we execute statements **5 - 8** of **if**. **5:**

$$(p_4^*, q_4^*) = (d \rightarrow a); \quad s_4^* = d; \quad t_4^* = a; \quad S_4 = C_4 = \{d\}.$$

**6:**

$$\Delta_4 = 0.5; \quad A_4 = a_{da}.$$

**7:**  $g_4^*$  contains the single arc  $(d \rightarrow a)$  (shown in blue in Fig. 5(c)). **8:** Set  $k = 3$  and return to the beginning of the main **while**-cycle.

**4:** The next minimal weight arc removed from  $M$  and added to  $G^*$  is  $(a \rightarrow b; 1)$ . Its endpoints  $a$  and  $b$  belong to different connected components of  $g_4^*$ . Hence we execute the statements of **if**. **5:**

$$(p_3^*, q_3^*) = (a \rightarrow b); \quad \Delta_3 = 1; \quad t_3^* = b, \quad S_3 = \{a, d\}; \quad C_3 = \{a\}.$$

**6:**

$$A_3 = a_{ab}; \quad s_3^* = a.$$

**7:**  $g_3^*$  has two arcs  $(d \rightarrow a; 0.5)$  and  $(a \rightarrow b; 1)$  shown in blue in Fig. 5(d). **8:** Set  $k = 2$  and return to the beginning of the main **while**-cycle.

**4:** The next minimal weight arc removed from  $M$  and added to  $G^*$  is  $(c \rightarrow a; 2)$ . Its endpoints belong to different connected components of  $g_3^*$ , hence we execute the statements of **if**. **5 :**

$$(p_2^*, q_2^*) = (c \rightarrow a); \quad s_2^* = c; \quad t_2^* = b; \quad S_2 = C_2 = \{c\}.$$

**6:**

$$\Delta_2 = 2; \quad A_2 = a_{ca}.$$

**7:**  $g_2^*$  has three arcs:  $(d \rightarrow a; 0.5)$ ,  $(a \rightarrow b; 1)$ , and  $(c \rightarrow a; 2)$  shown in blue in Fig. 5(e). **8:** Set  $k = 1$  and return to the beginning of the main **while**-cycle.

**4:** The next minimal weight arc removed from  $M$  and added to  $G^*$  is  $(b \rightarrow c; 3)$ . Its endpoints belong to the same connected component of  $g_2^*$ . Hence we execute the statements of **else**. **9:** We detect the cycle

$$C = \{(b \rightarrow c), (c \rightarrow a), (a \rightarrow b)\}$$

shown in Fig. 5(f) in red. **10:** Then we increase the arc weights in the sets  $B_a$  and  $B_c$  using the update rule (25) by  $3 - 1 = 2$  and  $3 - 2 = 1$  respectively:

$$\begin{aligned} B_a &= \{(a \rightarrow d; 3.8), (a \rightarrow e; 4), (a \rightarrow c; 7)\}; \\ B_c &= \{(c \rightarrow b; 3.6), (c \rightarrow d; 4.2)\}. \end{aligned}$$

The updated arc weights are shown in Fig. 5(g) in red. The pre-factors of arcs in  $B_a$  and  $B_c$  are multiplied by  $a_{bc}/a_{ab}$  and  $a_{bc}/a_{ca}$  respectively. **11:** Merge the sets  $B_a$ ,  $B_b$  and  $B_c$  and sort the arcs in the resulting set:

$$B_{abc} = \{(c \rightarrow b; 3.6), (a \rightarrow d; 3.8), (a \rightarrow e; 4), (c \rightarrow d; 4.2), (a \rightarrow c; 7), (b \rightarrow a; 10)\}.$$

Define the cycle  $C' = C(a) \cup C(b) \cup C(c) = \{a, b, c\}$ . **12:** Discard the minimal weight arc  $(c \rightarrow b; 3.6)$  from  $B_{abc}$  because its endpoints belong to the same cycle  $C'$ . **13:** Set  $B_a = B_b = B_c = B_{abc}$  and  $C(a) = C(b) = C(c) = C' = \{a, b, c\}$ . **14:** The next minimal weight arc in  $B_{abc}$  is  $(a \rightarrow d; 3.8)$ . Remove it from  $B_{abc}$  and add to  $M$  (Fig. 5(h)). At this point,  $M$  is:

$$M = \{(a \rightarrow d; 3.8), (e \rightarrow d; 6)\}.$$

Now, we return to the beginning of the main **while**-cycle.

**4:** Remove the minimal weight arc  $(d \rightarrow a; 3.8)$  from  $M$  and add it to  $G^*$ . Its endpoints  $d$  and  $a$  belong to the same connected component of  $g_2^*$ . Hence we execute the statements of **else**. We detect the cycle (Fig. 5(i))

$$C = \{(a \rightarrow d), (d \rightarrow a)\}.$$

**10.** According to the update rule (25), the arc weights in the set  $B_d$  are increased by  $3.8 - 0.5 = 3.3$ :

$$B_d = \{(d \rightarrow c; 4.8), (d \rightarrow e; 6.3)\},$$

and their pre-factors are multiplied by the factor  $a_{ad}/a_{da}$ . The updated weights are shown in Fig. 5(j) in red. **11:** Merge the sets  $B_d$  and  $B_{abc}$  and sort the arcs in the resulting set:

$$B_{abcd} = \{(a \rightarrow e; 4), (c \rightarrow d; 4.2), (d \rightarrow c; 4.8), (d \rightarrow e; 6.3), (a \rightarrow c; 7), (b \rightarrow a; 10)\}.$$

Define the cycle  $C' = C(a) \cup C(d) = \{a, b, c\} \cup \{d\} = \{a, b, c, d\}$ . The endpoint  $e$  of the minimal weight arc in  $B_{abcd}$  ( $a \rightarrow e; 4$ ) does not belong to  $C'$ , hence we skip **12**. **13:** Set  $B_i = B_{abcd}$ ,  $\min\_arc(i) = (a \rightarrow e)$ , and  $C(i) = \{a, b, c, d\}$  for  $i = a, b, c, d$ . **14:** The minimal weight arc  $(a \rightarrow e; 4)$  is removed from  $B_{abcd}$  and added to  $M$ . Now, we return to the beginning of the main **while**-cycle.

**4:** We remove the minimal weight arc  $(a \rightarrow e; 4)$  from  $M$  and add it to  $G^*$  (Fig. 5(k)). Its endpoints  $a$  and  $e$  belong to different connected components of  $g_2^*$ , hence we execute the statements of **if**. **5:**

$$(p_1^*, q_1^*) = (a \rightarrow e); \quad s_1^* = b; \quad t_1^* = e, \quad S_1 = C_1 = \{a, b, c, d\}.$$

**6:**

$$\Delta_1 = 4; \quad A_1 = \frac{a_{ae}a_{bc}}{a_{ab}}.$$

**7:** The optimal W-graph  $g_1^*$  is found by tracing the arcs from the set  $G^*$  backwards starting from the sink  $e = t_1^*$  so that each encountered vertex has only one outgoing arc (Fig. 5(l)). The arcs in  $G^*$  are blue and red. The arcs in  $g_1^*$  are blue. **8:** Set  $k = 0$ . Now,  $|M| = |\{(e \rightarrow d; 6)\}| = 1$ . Therefore, the main **while**-cycle is terminated. The found optimal W-graphs  $g_4^* - g_1^*$  are shown in Figs. 5(m) - (p) respectively.

From the output of Algorithm 1, we obtain the following asymptotic approximations for nonzero eigenvalues:

$$\lambda_4 \approx a_{da}e^{-0.5/T}, \quad \lambda_3 \approx a_{ab}e^{-1/T}, \quad \lambda_2 \approx a_{ca}e^{-2/T}, \quad \lambda_1 \approx \frac{a_{ae}a_{bc}}{a_{ab}}e^{-4/T}. \quad (26)$$

## 6.4 Implementation, cost, and performance

We have implemented Algorithm 1 both in Matlab and in C. Our C code uses pointers and such data structures as binary trees and link lists. All arcs are stored as structures containing three pointers to structures of the same type: parent, left child, and right child. The arcs in the sets  $B_i$ ,  $i \in S$ , and  $M$  are organized in binary trees automatically sorted using the heap sort [1]. As a result, removing or adding an arc from/to any of these sets is reduced to reshuffling the corresponding pointers to parents and children. Link lists are used to keep track of connected components of optimal W-graphs and Freidlin's cycles in a similar manner as in the description of Kruskal's algorithm in [1].



The computational cost of Algorithm 1 depends on the number vertices and the distributions of arcs and their weights. Let a network have  $n$  vertices and the maximal out-degree be  $d$ . The cost of the initialization is due to building  $n$  binary trees  $B_i$ , the removal of their roots, and building the main binary tree  $M$  out of them. It adds up to

$$O(nd \log d) + O(n \log d) + O(n \log n). \quad (27)$$

The computational cost of the main while-cycle depends on the complexity on the distribution of arcs and their weights. In the best case scenario, i.e., when no cycle is created and hence the else-statements are never executed, the cost is merely due to arc removals from the main tree  $M$ . It is

$$O(n \log n). \quad (28)$$

In the worst case scenario,  $n - 1$  cycles can be created. Then the computational cost will be dominated by the cost of merging the trees  $B_i$ . It will be at most

$$\sum_{l=d}^{nd} O(l \log l) = O((nd)^2 \log(nd)). \quad (29)$$

The performance of Algorithm 1 has been tested on the networks LJ<sub>38</sub> and LJ<sub>75</sub> created by Wales' group [45]. The LJ<sub>38</sub> network [44] contains 71,887 vertices and 239,706 arcs. The CPU time of Algorithm 1 applied to LJ<sub>38</sub> on a single core of an iMac desktop is 30 seconds. The number of created cycles is 50,226. The LJ<sub>75</sub> network contains 169,523 vertices and 441,016 arcs. The CPU time of Algorithm 1 applied to LJ<sub>75</sub> is 632 seconds. The number of created cycles is 153,164.

## 7 An application to LJ<sub>75</sub>

In this Section, we will present our analysis of the LJ<sub>75</sub> network. To make sure, it is time-reversible. For the purposes of the spectral analysis, we will restrict our attention to the largest connected component of LJ<sub>75</sub> containing 169,523 vertices. We will refer to the vertices by the indices of the corresponding local minima in the full dataset of 593,320 minima. A detailed description of Lennard-Jones clusters and, in particular, of LJ<sub>75</sub>, can be found in [43, 48]. Here we just remind that the energy landscape of LJ<sub>75</sub> has a double-funnel structure. Minimum 1 (vertex 1) in the LJ<sub>75</sub> network has the lowest energy and lies at the bottom of the deep and narrow funnel. It corresponds to the configuration shown in Fig. 1(a), the Marks decahedron with the point group  $D_{5h}$ . There are several local minima at the bottom of the wide and shallower icosahedral funnel. The one with the second lowest energy is minimum 92 (vertex 92) shown in Fig. 1(b).

The pairwise transition rates in the networks representing energy landscapes of Lennard-Jones clusters can be modeled using a harmonic approximation as it is done in [42]. In this case, the off-diagonal entries of the generator matrix are

$$L_{ij} = \frac{O_i (\Pi_+ \nu_i)^{1/2}}{O_{ij} (\Pi_+ \nu_{ij})^{1/2}} e^{-(V_{ij} - V_i)/T}, \quad (30)$$

where the subscripts  $i$  and  $ij$  refer to the minimum  $i$  and the transition state  $(ij)$  separating the minima  $i$  and  $j$  respectively.  $O_i$  and  $O_{ij}$  are the point group orders of  $i$  and  $(ij)$ ,  $\Pi_+ \nu_i$  and  $\Pi_+ \nu_{ij}$  are the products of the positive eigenvalues of the Hessian matrices at  $i$  and  $(ij)$ , and  $V_i$  and  $V_{ij}$  are the values of the potential at  $i$  and  $(ij)$  respectively. The invariant probability distribution is given by

$$\pi_i = \frac{1}{Z} \frac{e^{-V_i/T}}{O_i (\Pi_+ \nu_i)^{1/2}}, \quad \text{where} \quad Z = \sum_i \frac{e^{-V_i/T}}{O_i (\Pi_+ \nu_i)^{1/2}}. \quad (31)$$

## 7.1 Zero-temperature asymptotic analysis

### 7.1.1 Asymptotic estimates for eigenvalues of $\text{LJ}_{75}$

Eq. (19) implies that the asymptotic estimates for the eigenvalues of generator matrices with off-diagonal entries of the form of Eq. (30) are given by

$$\lambda_k = \frac{O_{s_k^*} (\Pi_+ \nu_{s_k^*})^{1/2}}{O_{p_k^* q_k^*} (\Pi_+ \nu_{p_k^* q_k^*})^{1/2}} e^{-(V_{p_k^* q_k^*} - V_{s_k^*})/T}. \quad (32)$$

The collection of numbers  $\Delta_k = V_{p_k^* q_k^*} - V_{s_k^*}$  for the  $\text{LJ}_{75}$  network is shown in Fig. 6. As one

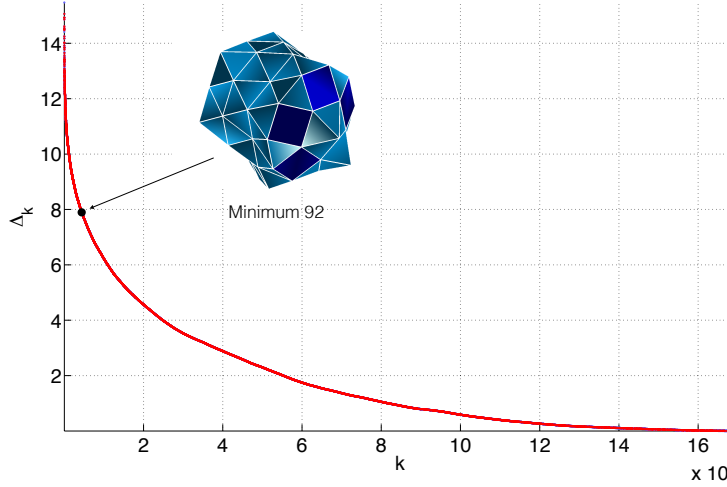


Figure 6: The exponential factors  $\Delta_k$ ,  $1 \leq k \leq 169522$ , determining the asymptotic estimates for the eigenvalues of the generator matrix of  $\text{LJ}_{75}$  according to Eq. (32).

can see, there are no notable gaps separating the exponential factors  $\Delta_k$ 's. Hence, there is no appreciable scale separation for the transition processes going on in  $\text{LJ}_{75}$ . As a result, one cannot find any collection of metastable points (a subset of states), with respect to which this Markov chain would be metastable according to the definition in [3, 4]. A similar phenomenon was observed in the  $\text{LJ}_{38}$  network [7].

The eigenvalue corresponding to the escape process from the icosahedral funnel is the one computed at that step of Algorithm 1 at which the vertex 92 ceases to be a sink. This happens at  $k = 4395$ . Therefore, minimum 92 is the sink  $s_{4395}^*$ . The corresponding sink  $t_{4395}^*$  is minimum 1. Hence, the escape from the icosahedral funnel brings the cluster into the Marks decahedron funnel with probability close to one at small enough temperatures. The pre-factor and exponent for  $k = 4395$  found by Algorithm 1 are  $A_{4395} = 147.2$  and  $\Delta_{4395} = 7.897$ . Hence, at low temperatures, the theoretical escape rate from the icosahedral funnel is given by

$$\lambda_{4392}(T) \approx 147.2 e^{-7.897/T}. \quad (33)$$

We will verify this prediction in Section 7.2. The arc  $p_{4395}^* \rightarrow q_{4395}^*$  is found to be (25811  $\rightarrow$  73992). The quasi-invariant set  $S_{4395}$  contains 92883 states.

### 7.1.2 The asymptotic zero-temperature path

The asymptotic zero-temperature path (the MinMax path) [6] connecting minima 92 and 1 can be extracted as the unique path from vertex  $s_{4395}^* = 92$  to vertex  $t_{4395}^* = 1$  in the optimal

W-graph  $g_{4395}^*$ . The potential energy along this path relative to the energy at minimum 1 is plotted in Fig. 7.

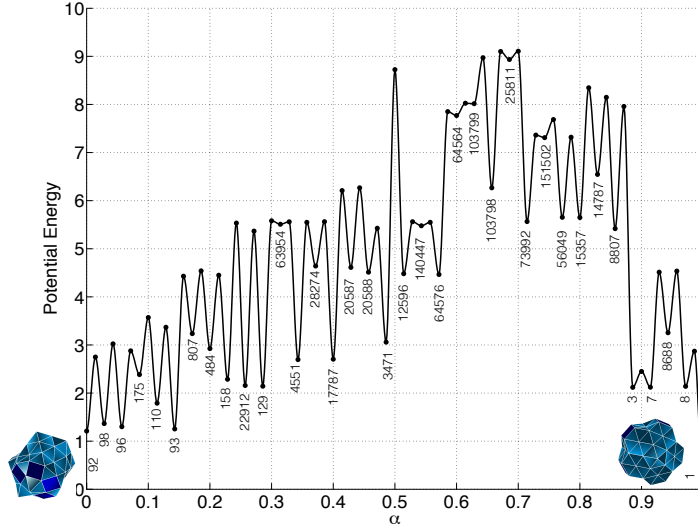


Figure 7: The energy along the asymptotic zero-temperature path (the MinMax path) connecting the two lowest minima 1 and 92 of LJ<sub>75</sub>. The energy at the global minimum (minimum 1) is assumed to be zero. The numbers next to each minimum indicate the indices of the corresponding local minima in D. Wales’s full dataset.

### 7.1.3 The decomposition in the maximal quasi-invariant sets

A quasi-invariant set  $S_k$  is called maximal if it is not a subset of any other quasi-invariant set  $S_l$  for  $0 < l < k$ . In other words, if  $t_k^* = s_0^*$ , i.e., the connected component with the sink  $s_k^*$  is absorbed by the connected component with sink  $s_0^*$  (the only sink of  $g_1^*$ ). The set of states of the LJ<sub>75</sub> network can be decomposed into the union of sink  $s_0^*$  (minimum 1) and 15692 maximal quasi-invariant sets. The largest maximal quasi-invariant set is  $S_{4395}$  with sink  $s_{4395}^* = 92$ . All maximal quasi-invariant sets with more than 1000 local minima are listed in Table 1. The numbers  $|C_k|$  are the numbers of local minima (vertices) in the sets  $C_k$  constructed by Algorithm 1. The sets  $C_k$  consist of all vertices separated from the sink  $s_k^*$  by energy barriers less than  $\Delta_k$ . They are the maximal Freidlin’s cycles containing minimum  $s_k^*$  and not containing any local minimum with energy less than  $V_{s_k^*}$  [17, 19, 18, 6, 7].

Table 1: The maximal quasi-invariant sets in LJ<sub>75</sub> containing over 1000 local minima.

$k$	$s_k^*$	$ S_k $	$ C_k $	$\Delta_k$
4395	92	92883	28032	7.897
47688	2	7141	27	2.418
71664	24	5026	52	1.308
30880	18438	2609	177	3.465
41622	135496	2590	38	2.785

A diagram showing the mutual arrangement of the maximal quasi-invariant sets containing

more than 100 local minima is shown in Fig. 8. Each shown maximal quasi-invariant set  $S_k$  is represented as a circle whose area is proportional to the number of minima in  $S_k$ . The black number next to each circle is the index of the local minimum corresponding to the sink  $s_k^*$ . The corresponding arcs ( $p_k^* \rightarrow q_k^*$ ) are shown with the black arrows. The dark red numbers next to them are the exponential factors  $\Delta_k$ .

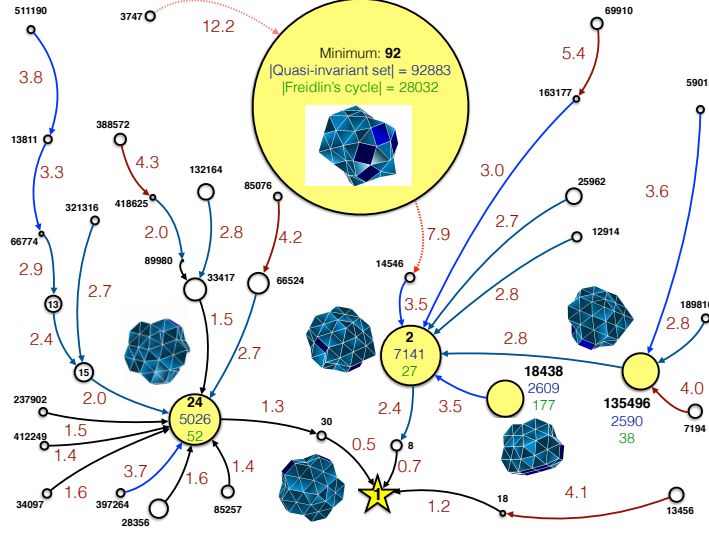


Figure 8: The maximal quasi-invariant sets in LJ<sub>75</sub> containing over 100 local minima.

#### 7.1.4 Using bond-order parameters $Q_4$ and $Q_6$

The set of local potential minima of LJ<sub>75</sub> can be mapped into a low-dimensional space of bond-order parameters introduced in [37, 38] in order to distinguish different types of atomic packings. The bond-order parameters are defined by

$$Q_l = \left[ \frac{4\pi}{2l+1} \sum_{m=-l}^l |\langle Y_{lm}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij})) \rangle|^2 \right]^{1/2},$$

where  $Y_{lm}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij}))$  are the spherical harmonics,  $\theta$  and  $\phi$  are the polar angles of the radius-vector of the bond  $\mathbf{r}_{ij}$ , and the average  $\langle Y_{lm}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij})) \rangle$  is taken over all bonds in the cluster. The bond-order parameters ( $Q_4, Q_6$ ) were used in e.g. [13, 31, 33, 25] in order to monitor structural transitions in Lennard-Jones clusters. Following [13], we set that atoms  $i$  and  $j$  in the cluster form a bond  $\mathbf{r}_{ij} := \mathbf{r}_i - \mathbf{r}_j$  if  $|\mathbf{r}_i - \mathbf{r}_j| < 1.391r^*$ , where  $\mathbf{r}_i$  and  $\mathbf{r}_j$  are the positions of atoms  $i$  and  $j$  respectively, and  $r^* = 2^{1/6}$  is the minimizer of the pair potential  $V(r) = 4(r^{-12} - r^{-6})$ .

In this Section, our goal is to investigate whether and how the bond-order parameters ( $Q_4, Q_6$ ) can be used for the study of the transition process between icosahedral and Marks decahedral configurations in LJ<sub>75</sub>. The set of states belonging to the quasi-invariant set  $S_{4395}$  with the sink  $s_{4395}^* = 92$  is shown with the light and dark blue dots in Fig. 9, while the rest of the states is shown with the pink and red dots. These sets of dots are all intermixed, hence it is impossible to separate them in the  $(Q_4, Q_6)$  plane. However, let us turn our attention to Freidlin's cycles. Freidlin's cycle  $C_{4395}$ , i.e., the set of local minima separated from minimum 92 by lower potential barriers than  $\Delta_{4395} = 7.897$ , is shown with dark blue dots in Fig. 9. It consists of 28032 local minima. The largest Freidlin's cycle containing minimum 1 and not containing minimum 92 consists of 15838 local minima. It is shown with red dots. These sets of

dots are intermixed only in a small region around  $Q_4 = 0.015$  and  $Q_6 = 0.17$ . We will refer to it as the transition region. The MinMax path is shown with the yellow piecewise linear curve. Comparing Figs. 7 and 9 we see that the MinMax path first wanders in the icosahedral corner ( $Q_4 < 0.01$ ,  $Q_6 < 0.03$ ) while passing through states 92 – 129. Next, it switches to the transition region passing through states 63954, 4551 and 2874. Then, it wanders in the transition region where it overcomes the highest potential barriers and passes through states 17787 – 8807. Finally, it jumps to state 3 lying in the Marks decahedral region ( $Q_4 \approx 0.03$  and  $Q_6 \approx 0.3$ ) and remains there while passing through states 7, 8688, 8, and reaching state 1. Hence, despite not perfect, the coordinates  $(Q_4, Q_6)$  can be used as reaction coordinates for the transition process between the two lowest minima of LJ<sub>75</sub>.

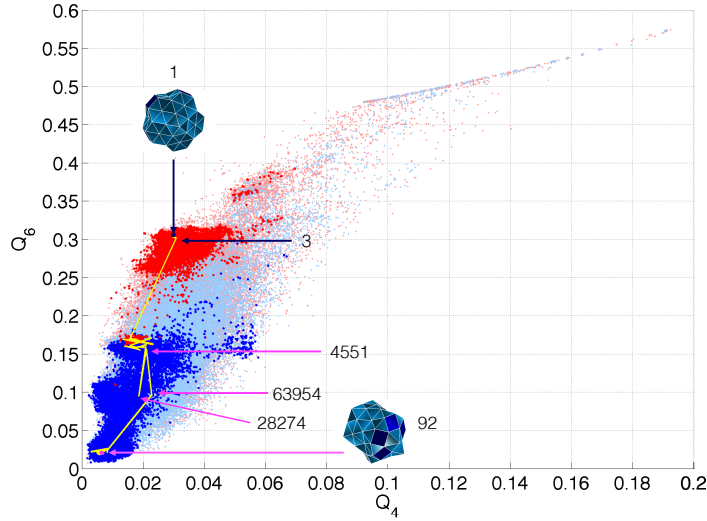


Figure 9: Some results of the asymptotic spectral analysis of LJ<sub>75</sub> presented using bond-order parameters  $(Q_4, Q_6)$ . Red dots: the largest Freidlin's cycle containing minimum 1 that does not contain minimum 92 (15838 local minima). Dark blue dots: the largest Freidlin's cycle containing minimum 92 that does not contain minimum 1 (28032 local minima). Light blue dots: the quasi-invariant set associated with minimum 92 (92883 local minima). Pink dots: the set of minima not contained in the quasi-invariant set associated with minimum 92 (76640 local minima). Piecewise linear yellow curve: the asymptotic zero-temperature path (the MinMax path).

## 7.2 Finite temperature analysis

In this Section, we describe our finite temperature continuation technique developed for computing the eigenvalue and the right eigenvector corresponding (at low temperatures) to the transition process from the icosahedral to the Marks decahedron funnel. Then we present a quantitative description of this transition process using the corresponding eigencurrent calculated from the right eigenvector. We will refer to these eigenvalue, right eigenvector and eigencurrent as  $\lambda(ICO - MARKS)$ ,  $\phi(ICO - MARKS)$ , and  $F(ICO - MARKS)$  respectively.

### 7.2.1 Continuation strategy

We would like to obtain the eigenpair  $(ICO - MARKS)$  at the temperature range  $0 < T \leq T_{sl}^{LJ75} = 0.25$  (the melting point). A similar task was accomplished for LJ<sub>38</sub> in [8]. However,

the continuation technique used for LJ<sub>38</sub> turned out to be unsuccessful for LJ<sub>75</sub>. The LJ<sub>75</sub> network is larger and more complex than LJ<sub>38</sub>. It contains wider range of potential barriers. The solid-solid transition temperature  $T_{ss} = 0.08$  [12] where Marks decahedral states give place to icosahedral configurations is extremely low (see Section 1) in comparison with the energetic barrier of 7.897 separating minimum 92 from minimum 1. For  $T < 0.17$ , Rayleigh quotient iterations (e.g. [40]) fail on LJ<sub>75</sub> due to issues associated with floating point arithmetic (produces NaN or fails to converge). For  $T \geq 0.17$ , the initial approximation for the right eigenvector (the indicator function of  $S_{4395}$ ) is not close enough to the right eigenvector  $\phi(ICO - MARKS)(T)$  at  $T \geq 0.17$ , so that Rayleigh quotient iterations converge to a wrong eigenpair. As a result, we have nothing to anchor on.

To tackle these difficulties, we deployed two remedies, lumping and truncation, and used a verification method relying on the eigencurrent. The resulting continuation strategy enabled us to obtain the eigenvalue  $\lambda(ICO - MARKS)$  for the temperature range  $0.05 \leq T \leq 0.25$  (see Fig. 11).

### 7.2.2 Initial approximation and verification

As we have mentioned in Section 2, if the continuous-time Markov chain is time-reversible, which is always the case if it represents an energy landscape, its generator matrix  $L$  is similar a symmetric matrix  $L_{sym} = P^{1/2}LP^{-1/2}$  where  $P = \text{diag}\{\pi_1, \dots, \pi_n\}$ . The matrix of eigenvectors of  $L_{sym}$  is orthogonal and given by  $P^{1/2}\Phi$ , where  $\Phi$  is the matrix of right eigenvectors of  $L$  normalized so that  $\Phi^T P \Phi = I$ . The Rayleigh quotient iteration is applied to the matrix  $L_{sym} = L_{sym}(T)$  (note that  $L_{sym}$  depends on temperature). Suppose we want to compute the eigenvector  $v(T)$  of  $L_{sym}(T)$  at temperatures  $T_1 < T_2 < \dots < T_m$  corresponding to the escape process described by the  $k$ th eigenpair  $(\lambda_k, \phi^k)$  of  $L$  at temperatures close to zero.

The initial approximation for  $v(T_l)$ ,  $1 \leq l \leq m$ , can be taken to be  $P^{1/2}(T_l)\phi_0^k$ , where  $\phi_0^k$  is the indicator function of  $S_k$  and  $P^{1/2}(T_l) = \text{diag}\{\pi_1^{1/2}(T_l), \dots, \pi_n^{1/2}(T_l)\}$ . Alternatively, suppose the Rayleigh quotient iteration has converged to the desired eigenpair some  $T_r$ . Then initial approximations for  $v(T_l)$ ,  $1 \leq l \leq m$ ,  $l \neq r$ , can be chosen to be  $P^{1/2}(T_l)P^{-1/2}(T_r)v(T_r)$ .

Suppose we have calculated an eigenpair  $(\lambda, v)$  of  $L_{sym}$  at temperature  $T$ . Then  $(\lambda, w = P^{-1/2}(T)v)$  is the eigenpair of  $L(T)$ . Now we need to verify whether it corresponds to the  $k$ th eigenpair  $(\lambda_k, \phi^k)$  of  $L$  at temperatures close to zero. We claim that  $(\lambda, w = P^{-1/2}(T)v)$  is the desired eigenpair if significant fractions of the corresponding eigencurrent are emitted at the state  $s_k^*$  and absorbed at the state  $t_k^*$ . Recall that the total amount of eigencurrent  $F$ , corresponding to the eigenpair  $(\lambda, w)$  of  $L$ , emitted or absorbed at state  $i$ , is given by

$$\sum_{j \neq i} F_{ij} = e^{-\lambda t} \lambda \pi_i w_i.$$

Therefore, if we sort the components for the vector  $P(T)w$  in the decreasing order, the component  $w_{s_k^*}$  must be the first or nearly the first, while the component  $w_{t_k^*}$  must be last or among the last ones.

### 7.2.3 Lumping and truncation

As we have mentioned, for  $T < 0.17$ , the Rayleigh quotient iteration fails to converge when we try to calculate the *ICO - MARKS* eigenpair, while for  $T \geq 0.17$ , the iterations converge to wrong eigenpairs. If we manage to obtain a better initial approximation to the desired eigenpair, we can hope to calculate it for  $0.17 \leq T \leq 0.25$ . At lower temperatures, only approximate eigenvalues can be calculated using a modified generator matrix  $\tilde{L}$  instead of  $L$ .

The Rayleigh quotient iteration fails to converge for  $T < 0.17$  due to large range of orders of magnitude of its entries. The largest entries of  $L$  can be eliminated by lumping together states in each connected component of the optimal W-graph  $g_k^*$  where  $k$  is the smallest number such that  $\Delta_k \leq \Delta_{\min}$ , a user-supplied threshold. The smallest entries of  $L$  can be removed

by capping the allowed values of the potential at states and edges by some user-supplied  $V_{\max}$ . Lumping and truncation are illustrated in Figs. 10 (a) and (b) respectively.

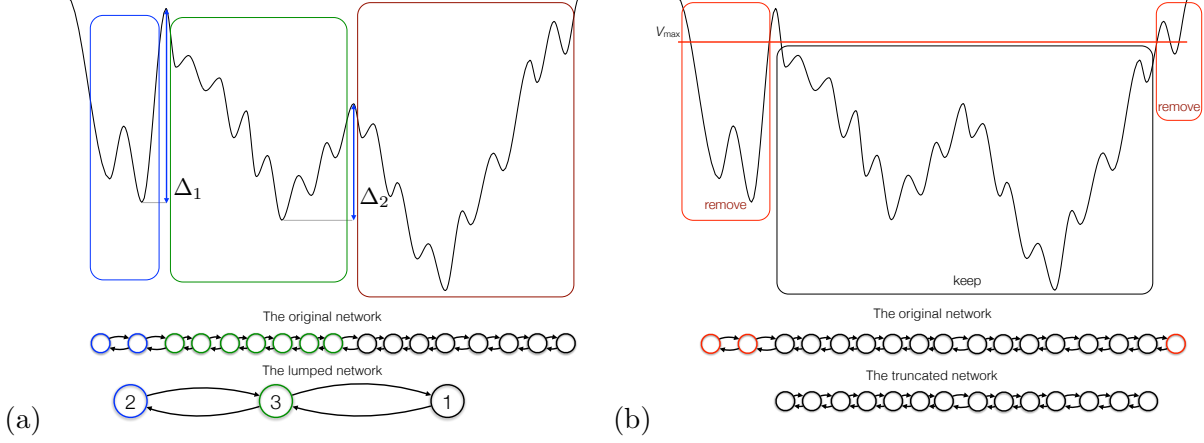


Figure 10: (a) Lumping of the network using  $\Delta_{\min} = \Delta_2$ . (b) Truncation of the network.

The pairwise transition rates in the lumped network are calculated as follows. Let the optimal W-graph  $g_k^*$  consist of connected components  $\Gamma_1, \dots, \Gamma_k$ . Then the lumped network has  $k$  states. The off-diagonal entries of its generator matrix  $\tilde{L}$  are defined according to the formula

$$\tilde{L}_{lr} = \sum_{i \in \Gamma_l} \frac{\pi_i}{\sum_{i' \in \Gamma_l} \pi_{i'}} \sum_{j \in \Gamma_r} L_{ij}, \quad l \neq r, \quad \tilde{L}_{ll} = -\sum_{r \neq l} \tilde{L}_{lr}, \quad 1 \leq j, r \leq k. \quad (34)$$

Eq. (34) says that the transition rate from the connected component  $\Gamma_l$  to the connected component  $\Gamma_r$  is the sum of transition rates  $L_{ij}$  along arcs  $i \rightarrow j$  where  $i \in \Gamma_l$  and  $j \in \Gamma_r$  multiplied by the conditional probabilities  $\pi_i / (\sum_{i' \in \Gamma_l} \pi_{i'})$  to find the system at state  $i$  provided that it is in  $\Gamma_l$ . The diagonal entries of  $\tilde{L}$  are defined so that its row sums are zero.

One can check [20, 21] that the  $k \times k$  generator matrix  $\tilde{L}$  of the lumped network is obtained from the  $n \times n$  generator matrix  $L$  of the original network by multiplying  $L$  by a  $k \times n$  matrix  $A$  from the left and by a  $n \times k$  matrix  $B$  from the right

$$\begin{matrix} \tilde{L} & = & A & L & B \\ k \times k & & k \times n & n \times n & n \times k \end{matrix} \quad (35)$$

where the matrices  $A$  and  $B$  are defined by

$$A_{li} = \begin{cases} \frac{\pi_i}{\sum_{i' \in \Gamma_l} \pi_{i'}}, & i \in \Gamma_l, \\ 0, & i \notin \Gamma_l, \end{cases} \quad B_{jr} = \begin{cases} 1, & j \in \Gamma_r, \\ 0, & j \notin \Gamma_r. \end{cases}, \quad 1 \leq i, j \leq n, \quad 1 \leq l, r \leq k. \quad (36)$$

In words,  $l$ -th row of  $A$  is the invariant distribution in  $\Gamma_l$  and zeros elsewhere, while  $r$ -th column of  $B$  is the indicator function of the connected component  $\Gamma_r$ . One can readily check that  $AB = I$ . Furthermore, if  $(\mu, x)$  is an eigenpair of  $\tilde{L}$ , then  $(\mu, Bx)$  is an eigenpair of  $B\tilde{L}A$ , which is a homogenized approximation to  $L$ . Therefore, once we have computed an eigenpair  $(\mu, x)$  of  $\tilde{L}$ , we can convert it to the corresponding eigenpair  $(\mu, Bx)$  of the  $n \times n$  matrix  $B\tilde{L}A$  and use  $Bx$  as the initial vector for the Rayleigh quotient iteration applied to the generator matrix  $L$  of the full network.

The truncation procedure consists of two steps. First, all edges  $(i, j)$  with  $V_{ij} > V_{\max}$  are removed from the network. Then the connected component of the network containing the global minimum  $s_0^*$  is extracted. If we need to combine truncation and lumping, we first truncate the

network and then lump it. Note that the combination of truncation and lumping allows us, in principle, to obtain an approximation to any desired eigenpair. Indeed, suppose we want to continue the asymptotic eigenpair  $(A_k \exp(-\Delta_k/T), \phi_0^k)$  to a range of finite temperatures. If we are unable to do it for the full network, we can pick

$$V_{\max} \geq \Delta_k + V_{s_k^*} \quad \text{and} \quad \Delta_{\min} \leq \Delta_k, \quad (37)$$

truncate and lump the network, and apply the Rayleigh Quotient iteration to the resulting symmetrized generator matrix with the appropriate initial approximation. If  $\Delta_{\min} = \Delta_k$  and  $V_{\max} = \Delta_k + V_{s_k^*}$ , the resulting generator matrix will be  $2 \times 2$ . Hence, its eigenvalue can be readily computed. However, it might be less accurate than we would like it to be. The farther  $\Delta_{\min}$  and  $V_{\max}$  from their bounds, the more challenging is the Rayleigh quotient iteration, but the better the computed eigenpair will approximate the exact one. Therefore, we pick  $V_{\max}$  as large as possible and  $\Delta_{\min}$  as small as possible so that the Rayleigh quotient iteration still converges.

### 7.2.4 The eigenvalue $\lambda(ICO - MARKS)$

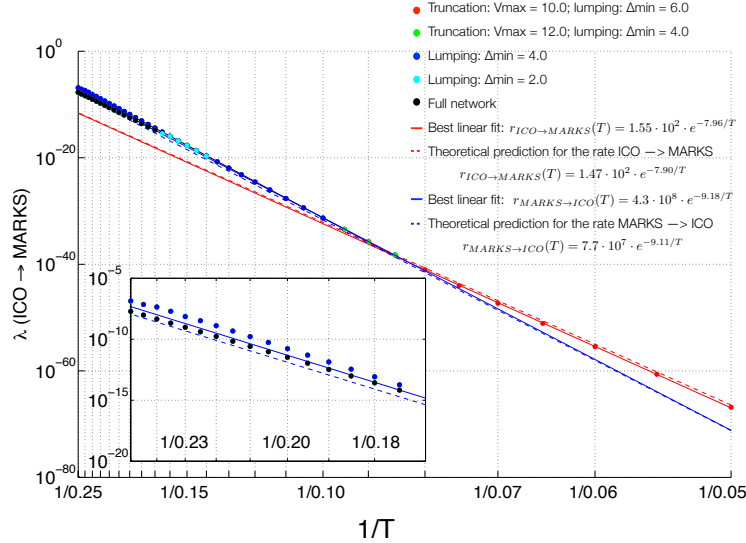


Figure 11: The eigenvalue  $ICO - MARKS$  corresponding to the transition process between the icosahedral and the Marks decahedral funnels in LJ<sub>75</sub>.

In this Section, we discuss the behavior of the computed eigenvalue  $\lambda(ICO - MARKS)(T)$  for  $0.05 \leq T \leq 0.25$  presented in Fig. 11. We remind that the asymptotic estimate for this eigenvalue as  $T \rightarrow 0$  is given by  $1.472 \cdot 10^2 \exp(-7.897/T)$ . We were able to compute:

- red dots:** an approximation to  $\lambda(ICO - MARKS)(T)$  using the truncated and lumped network with  $V_{\max} = 10.0$  and  $\Delta_{\min} = 6.0$  for  $0.05 \leq T \leq 0.25$ ;
- green dots:** an approximation to  $\lambda(ICO - MARKS)(T)$  using the truncated and lumped network with  $V_{\max} = 12.0$  and  $\Delta_{\min} = 4.0$  for  $0.085 \leq T \leq 0.25$ ;
- blue dots:** an approximation to  $\lambda(ICO - MARKS)(T)$  using the lumped network with  $\Delta_{\min} = 4.0$  for  $0.10 \leq T \leq 0.25$ ;
- cyan dots:** an approximation to  $\lambda(ICO - MARKS)(T)$  using the lumped network with  $\Delta_{\min} = 2.0$  for  $0.14 \leq T \leq 0.25$ ;



**black dots:**  $\lambda(ICO - MARKS)(T)$  using the full network for  $0.17 \leq T \leq 0.25$ . The initial approximations for the eigenvectors were obtained from the lumped network with  $\Delta_{\min} = 2.0$  as described in Section 7.2.3.

The numbers of vertices  $N$  and maximal vertex degrees  $d_{\max}$  of the corresponding networks are listed in Table 2.

Table 2: The number of vertices and maximal vertex degrees in the truncated and lumped network.  $V_{\max} = \infty$  means no truncation.  $\Delta_{\min} = 0.0$  means no lumping. The last row corresponds to the full LJ<sub>75</sub> network.

$V_{\max}$	$\Delta_{\min}$	$N$	$d_{\max}$
10.0	6.0	91	90
12.0	4.0	9361	6745
$\infty$	4.0	24833	13888
$\infty$	2.0	55251	6663
$\infty$	0.0	169523	740

The blue, green, and red dots in Fig. 11 visually coincide in the intersection of their intervals  $0.10 \leq T \leq 0.25$ , while the black and cyan dots visually coincide on the intersection of their intervals  $0.17 \leq T \leq 25$ . The inset zooms in the plots at the temperature range  $0.17 \leq T \leq 25$ . The cyan and black dots are indistinguishable in it too. Beside that, the blue and the cyan dots are indistinguishable for  $0.14 \leq T \leq 0.16$ .

The graph of the estimate of the eigenvalue  $\lambda(ICO - MARKS)(T)$  versus  $1/T$  in the logarithmic scale in the  $y$ -coordinate is nearly a piecewise linear curve consisting of two line segments (Fig. 11). The best least squares fit with a linear function on the interval  $0.05 \leq T \leq 0.08$  (the solid red line) can be compared to the asymptotic estimate obtained by Algorithm 1 (the dashed red line):

$$\begin{aligned} \text{Algorithm 1 : } \lambda_{4395} &\equiv \lambda(ICO \rightarrow MARKS) \approx 1.472 \cdot 10^2 e^{-7.897/T}, \\ \text{best linear fit : } \lambda(ICO - MARKS) &\approx 1.552 \cdot 10^2 e^{-7.961/T}. \end{aligned} \quad (38)$$

For  $T > 0.08$ , the best least squares fit with a linear function (the solid blue line) (obtained for  $0.08 \leq T \leq 0.12$ ) can be compared to the asymptotic estimate for the eigenvalue  $\lambda(MARKS \rightarrow ICO)$  obtained pretending that the deepest minimum is located in the icosahedral funnel:

$$\begin{aligned} \text{Prediction : } \lambda(MARKS \rightarrow ICO) &\approx 7.674 \cdot 10^7 e^{-9.107/T}, \\ \text{best linear fit : } \lambda(ICO - MARKS) &\approx 4.307 \cdot 10^8 e^{-9.187/T}. \end{aligned} \quad (39)$$

Recall that  $T_{ss}^{LJ75} = 0.08$  is the solid-solid transition critical temperature. For  $T < 0.08$ , the system spends more time in the Marks decahedral funnel rather than in the icosahedral one, while for  $T > 0.08$  it is the other way around. This critical temperature is extremely low in comparison with  $\Delta(ICO - MARKS) = 7.897$ . As Fig. 11 demonstrates, at  $T < 0.08$ , the asymptotic estimate for the eigenvalue is close to the estimate computed using truncated and lumped matrix with  $V_{\max} = 10.0$  and  $\Delta_{\min} = 6.0$ . For  $T$  slightly larger than  $T = 0.08$ , the system spends more time in the icosahedral funnel with several local minima with close values of the potential at its bottom (minima 92, 93, 94, 95, 96, 97, 98, ...). Effectively, they can be lumped together into one minimum with the value of the potential lower than that at minimum 1. This is why the predicted asymptotic estimate (39) turns out to be quite accurate. Due to the fact that the barrier  $\Delta = 9.107$  separating state 1 from the icosahedral funnel is very high in comparison with  $T_{sl}^{LJ75} = 0.25$ , this prediction remains quite accurate up to  $T_{sl}^{LJ75} = 0.25$  as shows the inset in Fig. 11.

### 7.2.5 The eigencurrent $F(ICO - MARKS)$

The computed eigenvector  $\phi(ICO - MARKS)$  allows us to quantify the transition process between the Marks decahedral and icosahedral funnels in terms of distributions of the corresponding eigencurrent  $F(ICO - MARKS)$ . Important characteristics of this process are the distributions of the emission and absorption of the eigencurrent, the location of the emission-absorption cut (the EA-cut), and the distribution of the eigencurrent in the EA-cut. Furthermore, if the eigencurrent happen to focus around some subset of edges, one can extract this set of edges and hence identify major reactive channel or channels.

First, we establish the location of the emission-absorption cut as a function of temperature. As the temperature increases, the location of the EA-cut along the MinMax path (Fig. 12)(a) shifts toward the icosahedral funnel. At  $T \rightarrow 0$ , it must be at the edge (25811, 73992) corresponding to the highest potential barrier separating states 1 and 92. Its height is 9.107 with respect to the potential at minimum 1. At  $0.17 \leq T \leq 0.235$  the emission-absorption cut intersects the MinMax path at the edge (3471, 12596) corresponding to another barrier separating states 1 and 92 that is almost as high as the highest one: its height is 8.724 with respect to minimum 1. At  $T = 0.24$ , the emission-absorption cut further drifts toward minimum 92 to the edge (20588, 3472), and reaches the edge (20587, 20588) at  $0.245 \leq T \leq 0.25$ . This shift can be explained by the entropic effect: the icosahedral funnel is significantly wider than the Marks decahedral one. As the system overcomes the highest barrier at relatively high temperature ( $T \geq 0.17$ ), it is more likely for it to overcome it back than to find its way to the bottom of the icosahedral funnel due to its width and complexity.

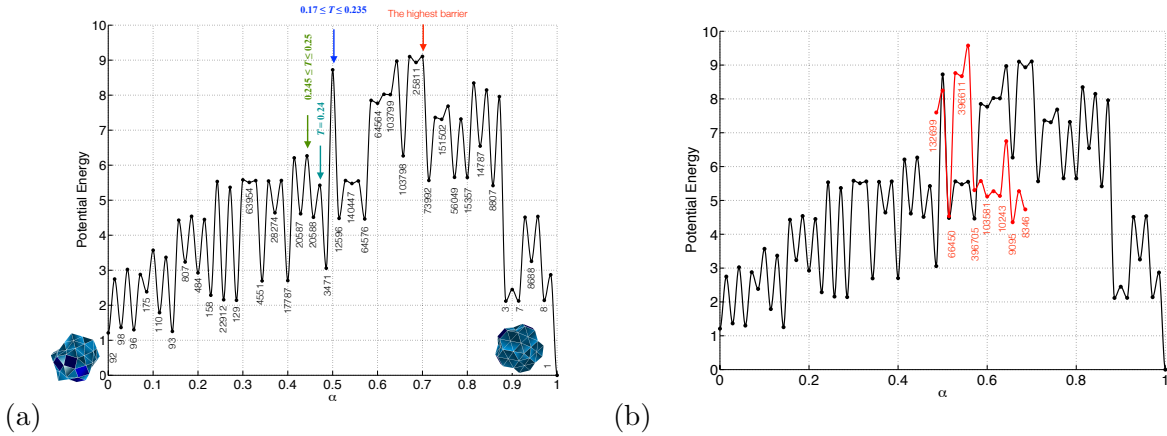


Figure 12: (a): The locations of the emission-absorption cut for the eigencurrent  $F(ICO - MARKS)$  along the MinMax path at temperatures  $T \rightarrow 0$  and  $0.17 \leq T \leq 0.25$ . (b): The energy plot (red) along the sequence of edges along which the eigencurrent  $F(ICO - MARKS)$  is focused at  $0.17 \leq T \leq 0.25$ .

Comparing the values of the eigencurrent  $F(ICO - MARKS)$  with its distribution in the EA-cut, we can extract a sequence of edges

$$132699 \rightarrow 66450 \rightarrow 396611 \rightarrow 396705 \rightarrow 103581 \rightarrow 10243 \rightarrow 9095 \rightarrow 8345,$$

along which the eigencurrent  $F(ICO - MARKS)$  is concentrated (see Fig. 13) at  $0.17 \leq T \leq 0.25$ . At  $T = 0.17$ , about 90% of the eigencurrent follows it. At  $T = 0.19$ , this number reduces to 76% and drops down to 36%  $T = 0.25$ . This sequence is not a subsequence of the asymptotic zero-temperature path. The energy plot along it is shown in Fig. 12(b). The EA-cut shifts toward the icosahedral funnel as the temperature increases. The eigencurrent widely spreads

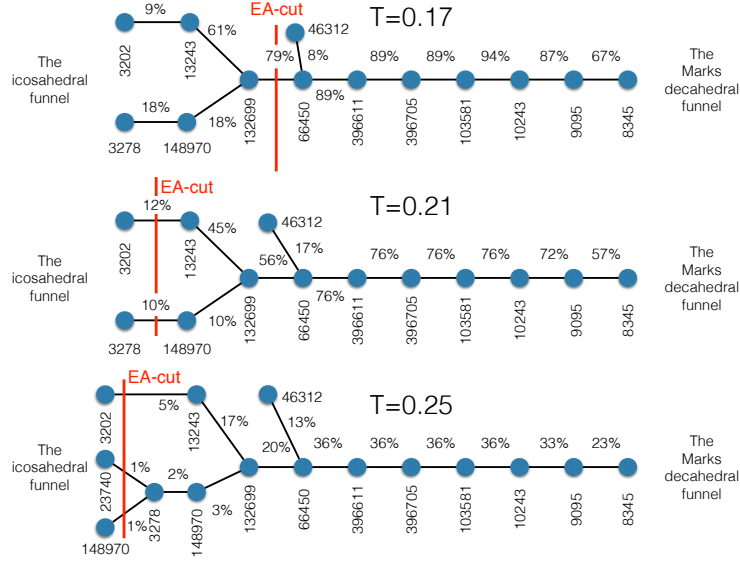


Figure 13: The network of edges passing through the EA-cut along which the eigencurrent  $F(ICO - MARKS)$  is concentrated.

out in the icosahedral funnel. So it does, but to a lesser extent, in the Marks decahedral funnel. The distributions of the eigencurrent  $F(ICO - MARKS)$  in the EA-cut at  $0.17 \leq T \leq 0.25$  are shown in Fig. 14(a). The eigencurrent is highly focused within the EA-cut  $0.17 \leq T \leq 0.19$ , while for  $0.21 \leq T \leq 0.25$ , it is spread out. This is caused by the shift of the EA-cut toward the icosahedral funnel as well as the spreading of the eigencurrent due temperature increase.

The emission (solid curves) and absorption (dashed curves) distributions of the eigencurrent at  $0.17 \leq T \leq 0.25$  are shown in Fig. 14(b). The intermediate asymptotics for these distributions in the log-log scale are straight lines, which means that they have heavy (power law) tails. While nearly all eigencurrent at  $T = 0.17$  is absorbed by two states: 1 and 239139, its emission is spread out across the icosahedral funnel. As the temperature increases, both distributions of emission and absorption widen (see Table 3). Fig. 15 visualizes the emission and absorption distributions in the plane of the bond-order parameters  $(Q_4, Q_6)$ . The emitting and absorbing states can be separated, though not perfectly, in the  $(Q_4, Q_6)$  - plane.

Table 3: The numbers of states emitting ( $N_e$ ) and absorbing ( $N_a$ ) 90% and 99% of the eigencurrent  $F(ICO - MARKS)$

Temperature	$N_e(0.90)$	$N_e(0.99)$	$N_a(0.90)$	$N_a(0.99)$
0.17	41	513	2	2
0.19	111	1437	2	6
0.21	270	2941	2	20
0.23	646	4741	4	145
0.25	1342	6737	16	862

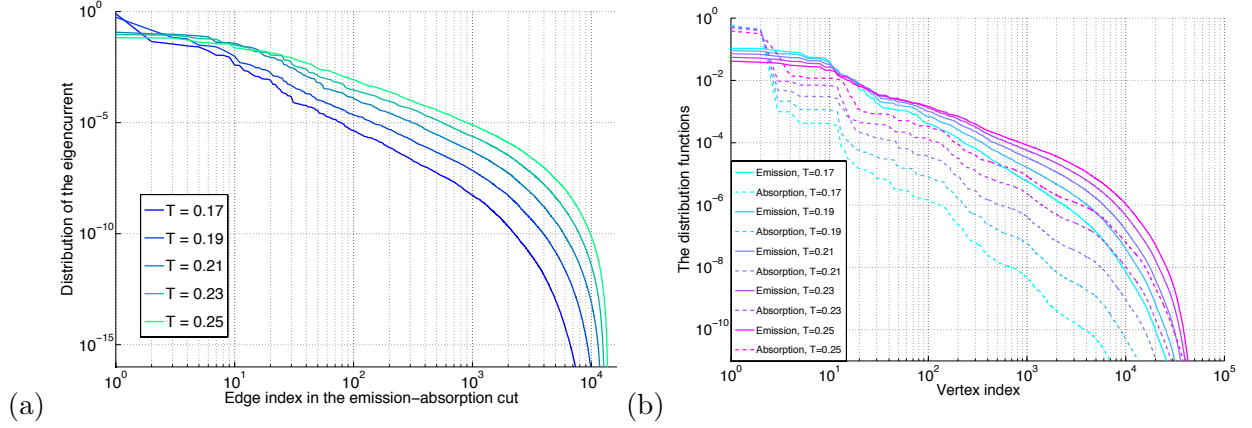


Figure 14: (a): The distributions of the eigencurrent  $F(ICO-MARKS)$  in the emission-absorption cut at  $0.17 \leq T \leq 0.25$ . (b): The emission (solid curves) and absorption (dashed curves) distributions of the eigencurrent  $F(ICO-MARKS)$   $0.17 \leq T \leq 0.25$ .

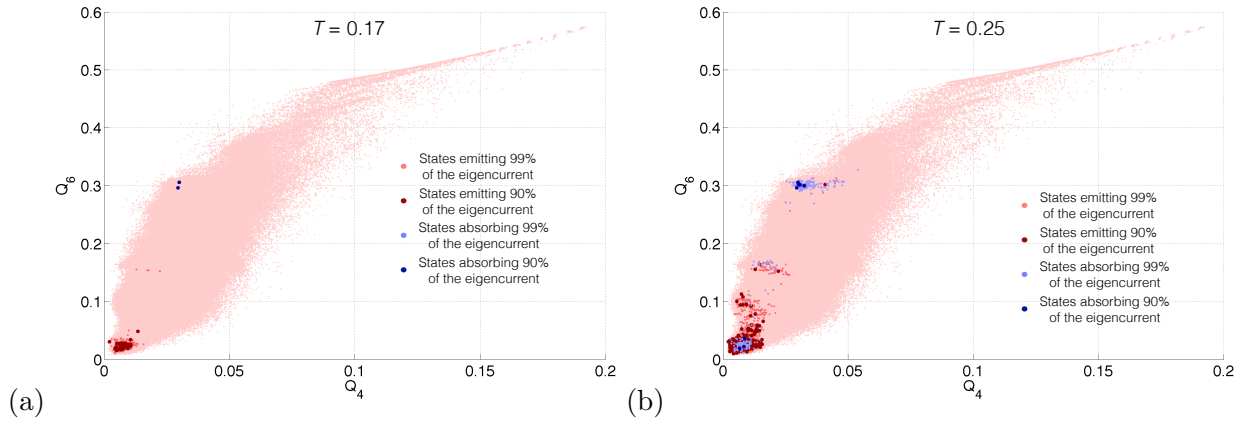


Figure 15: (a): States emitting and absorbing most of the eigencurrent  $F(ICO-MARKS)$  at  $T = 0.17$  in the bond-order parameters  $Q_4$  and  $Q_6$ . (b): States emitting and absorbing most the eigencurrent  $F(ICO-MARKS)$  at  $T = 0.25$  in the bond-order parameters  $Q_4$  and  $Q_6$ .

## 8 Concluding remarks

We have presented the two - step approach for spectral analysis of large stochastic networks. Step one is the asymptotic analysis and step two is the finite temperature continuation. Its power is demonstrated on the large and complex stochastic network representing the energy landscape of LJ<sub>75</sub>.

The theoretical justification for the asymptotic part of this approach is completed for time-reversible networks. For time-irreversible networks, we have extended Wentzell's result [49, 18] obtained sharp estimates for the eigenvalues (Eqs. (13)-(15)), i.e., containing both exponents and pre-factors. Algorithm 1 is equally efficient for both time-reversible and time-irreversible continuous-time Markov chains. We have proven the nested properties of the optimal W-graphs (Theorem 1) justifying the construction implemented in Algorithm 1. We are planning to address the problem of obtaining asymptotic estimates for left and right eigenvectors in the future.

The finite temperature continuation technique used for LJ<sub>75</sub> is quite general and can be used for an arbitrary time-reversible network of the considered type. The combination of truncation and lumping can fight floating point arithmetic issues and enable us to obtain approximations to eigenvalues and eigenvectors. These approximations for eigenvectors, in turn, can be used for obtaining initial guesses for eigenvectors for the Rayleigh quotient iteration applied to a less lumped and less truncated network. And so on, until we are able to compute eigenvalues and eigenvectors for the full network.

The continuation technique is yet to be extended to the time-irreversible case where the generator matrix is not symmetrizable and eigenpairs might become complex .

Our analysis of LJ<sub>75</sub> can be compared to the one of LJ<sub>38</sub> [6, 9, 7, 8]. Both of these networks represent energy landscapes with two major funnels one of which is deep and narrow and the other one is wide and not as deep. Both of these networks are large and complex and there is no scale separation between various transition processes in them. This causes the observed absence of spectral gaps (compare Fig. 6 to Fig. 6 in [7]). For both of these networks, the transition processes between two major funnels become diverse as the temperature increases, and the distributions of the corresponding eigencurrents in the corresponding EA-cut look similar (compare Fig. 14(a) to Fig. 8 in [8]). For both of these distributions, the intermediate asymptotics are power laws.

However, there is an important difference in the qualitative behavior of the eigenvalues corresponding to the transition process between the two major funnels of LJ<sub>75</sub> and LJ<sub>38</sub>. It is caused by the fact that the solid-solid phase transition temperature in LJ<sub>75</sub> is lower than that in LJ<sub>38</sub>, while the potential barrier separating the two funnels in LJ<sub>75</sub> is approximately twice as high as that in LJ<sub>38</sub>. In LJ<sub>75</sub>, the graph of the eigenvalue  $\lambda(ICO - MARKS)$  versus  $1/T$  in the logarithmic scale is a piecewise linear curve with a kink at  $T_{ss}^{LJ75}$  (Fig. 11) and linear pieces nearly perfectly agreeing with the theoretical estimates. In LJ<sub>38</sub>, the graph of the eigenvalue  $\lambda(ICO - FCC)$  versus  $1/T$  in the logarithmic scale is a smooth curve, because the transition process dramatically broadens before  $T_{ss}^{LJ38}$  is reached (Fig. 7 in [8]).

## 9 Acknowledgements

We thank Professor David Wales for providing us with the data for the LJ<sub>75</sub> network and valuable discussion. We are grateful to Mr. Weilin Li for valuable discussions at the early stages of the development of Algorithm 1. This work was partially supported by NSF grant 1217118.

## APPENDIX

**Proof of Theorem 1.** For brevity, we will denote the set of arcs with tails belonging to a subset of states  $\tilde{S}$  in the W-graph  $g$  by  $\mathcal{A}(g; \tilde{S})$ , and the sum of weights of these arcs by  $\Sigma(g; \tilde{S})$ ,

i.e.,

$$\mathcal{A}(g; \tilde{S}) := \{(i \rightarrow j) \in g \mid i \in \tilde{S}\}, \quad \Sigma(g; \tilde{S}) := \sum_{\mathcal{A}(g; \tilde{S})} U_{ij}.$$

*Proof.* First we prove Claims 1 and 2. Since there are  $k$  sinks in  $g_k^*$  and  $k+1$  connected components in  $g_{k+1}^*$ , at least one connected component of  $g_{k+1}^*$  contains no sink of  $g_k^*$ . Pick one such connected component of  $g_{k+1}^*$  and denote its set of vertices and its sink by  $S_k$  and  $s_k^*$  respectively.

The subset of vertices  $S \setminus S_k$  contains  $k$  sinks of  $g_{k+1}^*$ . Hence

$$|\mathcal{A}(g_{k+1}^*; S \setminus S_k)| = |S \setminus S_k| - k.$$

Since  $S_k$  contains no sinks of  $g_k^*$ , all  $k$  sinks of  $g_k^*$  lie in  $S \setminus S_k$ . Hence

$$|\mathcal{A}(g_k^*; S \setminus S_k)| = |S \setminus S_k| - k = |\mathcal{A}(g_{k+1}^*; S \setminus S_k)|.$$

Since the W-graphs  $g_{k+1}^*$  and  $g_k^*$  are optimal, we have that

$$\Sigma(g_{k+1}^*; S \setminus S_k) = \Sigma(g_k^*; S \setminus S_k) = \min_{g \in G(k; S \setminus S_k)} \sum_{(i \rightarrow j) \in g} U_{ij},$$

where  $G(k; S \setminus S_k)$  denotes the set of all W-graphs with  $k$  sinks defined on the set of vertices  $S \setminus S_k$ . Since the optimal W-graphs  $g_{k+1}^*$  and  $g_k^*$  are unique, we conclude that

$$\mathcal{A}(g_k^*; S \setminus S_k) = \mathcal{A}(g_{k+1}^*; S \setminus S_k).$$

This implies that the sinks in  $S \setminus S_k$  of  $g_k^*$  and  $g_{k+1}^*$  coincide. Hence, the connected component  $S_k$  of  $g_{k+1}^*$  containing no sinks of  $g_k^*$  is unique and every sink of  $g_k^*$  is also a sink of  $g_{k+1}^*$ . Thus, Claims 1 and 2 are proven.

Now we prove Claim 3. Suppose that there are  $l > 1$  arcs with tails in  $S_k$  and heads in  $S \setminus S_k$  in  $g_k^*$ . Then the set of vertices  $S_k$  can be subdivided into two subsets,  $\tilde{S}_k$  and  $S_k \setminus \tilde{S}_k$ , so that  $\tilde{S}_k$  is connected in  $g_k^*$ , contains the sink  $s_k^*$ , and  $S_k \setminus \tilde{S}_k$  and  $\tilde{S}_k$  are not connected in  $g_k^*$ , as shown in Fig. 16.

By construction,  $S_k \setminus \tilde{S}_k$  contains no sinks of  $g_{k+1}^*$ , hence

$$|\mathcal{A}(g_{k+1}^*; S_k \setminus \tilde{S}_k)| = |S_k \setminus \tilde{S}_k|,$$

i.e., every vertex in  $S_k \setminus \tilde{S}_k$  has an outgoing arc. Since all directed paths in  $g_{k+1}^*$  starting in the vertices of  $S_k \setminus \tilde{S}_k$  lead to the sink  $s_k^*$ ,

$$\mathcal{A}(g_{k+1}^*; S_k \setminus \tilde{S}_k) \neq \mathcal{A}(g_k^*; S_k \setminus \tilde{S}_k).$$

However, the set  $S_k \setminus \tilde{S}_k \subset S_k$  contains no sinks of  $g_k^*$ . Hence

$$|\mathcal{A}(g_k^*; S_k \setminus \tilde{S}_k)| = |S_k \setminus \tilde{S}_k| = |\mathcal{A}(g_{k+1}^*; S_k \setminus \tilde{S}_k)|.$$

Suppose we replace  $\mathcal{A}(g_{k+1}^*; S_k \setminus \tilde{S}_k)$  with  $\mathcal{A}(g_k^*; S_k \setminus \tilde{S}_k)$  in  $g_{k+1}^*$ . We create no cycles by this replacement, because (i) there is no arc from  $S_k \setminus \tilde{S}_k$  to  $\tilde{S}_k$  in  $g_k^*$ , and (ii) there is no arc from  $S \setminus S_k$  to  $S_k$  in  $g_{k+1}^*$ . Since  $g_{k+1}^*$  is the unique optimal W-graph with  $k+1$  sinks, we have:

$$\Sigma(g_{k+1}^*; S_k \setminus \tilde{S}_k) < \Sigma(g_k^*; S_k \setminus \tilde{S}_k). \quad (\text{A-1})$$

On the other hand, suppose we replace  $\mathcal{A}(g_k^*; S_k \setminus \tilde{S}_k)$  with  $\mathcal{A}(g_{k+1}^*; S_k \setminus \tilde{S}_k)$  in  $g_k^*$ . Since there is no arc from  $\tilde{S}_k$  to  $S_k \setminus \tilde{S}_k$  in  $g_k^*$ , this replacement creates no cycle. Since  $g_k^*$  is the unique optimal W-graph with  $k$  sinks, we have:

$$\Sigma(g_k^*; S_k \setminus \tilde{S}_k) > \Sigma(g_{k+1}^*; S_k \setminus \tilde{S}_k). \quad (\text{A-2})$$

Eq. (A-1) contradicts to Eq. (A-2). Hence the assumption that there is more than one outgoing arc from  $S_k$  to  $S \setminus S_k$  in  $g_k^*$  leads to a contradiction. This proves Claim 3.  $\square$

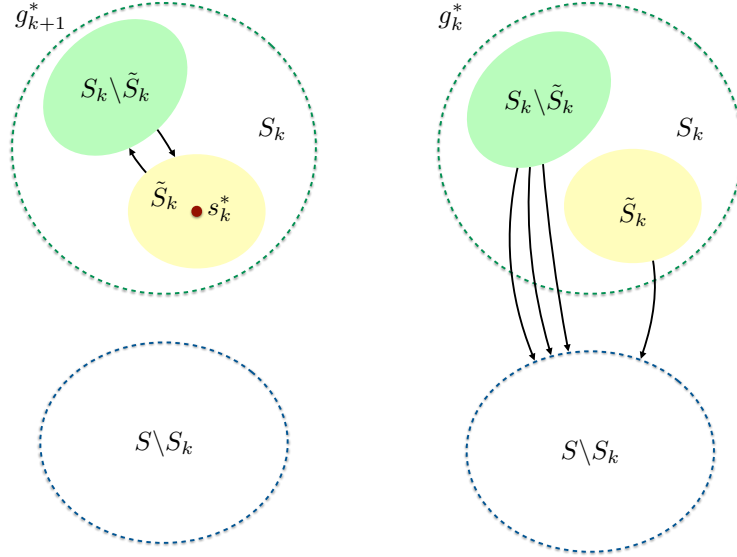


Figure 16: An illustration for the proof of Claim 3 of Theorem 1.

## References

- [1] R. K. Ahuja, T. L. Magnanti, J. B. Orlin, “Network flows: Theory, Algorithms, and Applications”, Prentice Hall, New Jersey, 1993.
- [2] R. D. Astumian, Biasing the random walk of a molecular motor, J. Phys.: Condensed Matter, **17** (2005) S3753-S3766
- [3] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, Metastability and Low Lying Spectra in Reversible Markov Chains, Comm. Math. Phys. **228** (2002), 219-255
- [4] A. Bovier, Metastability, in “Methods of Contemporary Statistical Mechanics”, (ed. R. Kotecky), LNM 1970, Springer, 2009
- [5] A. Bovier and F. den Hollander, “Metastability: A Potential-Theoretic Approach”, Springer, 2016
- [6] M. K. Cameron, Computing Freidlin’s cycles for the overdamped Langevin dynamics, J. Stat. Phys. **152**, 3 (2013), 493-518
- [7] M. Cameron, Computing the asymptotic spectrum for networks representing energy landscapes using the minimum spanning tree, Networks and Heterogeneous Media, **9**, 3 (2014), 383-416
- [8] M. Cameron, Metastability, Spectrum, and Eigencurrents of the Lennard-Jones-38 Network, J. Chem. Phys. **141**, (2014) 184113
- [9] M. K. Cameron and E. Vanden-Eijnden, Flows in Complex Networks: Theory, Algorithms, and Application to Lennard-Jones Cluster Rearrangement, J. Stat. Phys., **156** (2014), 427
- [10] J. D. Chodera and N. Singhal and V. S. Pande and K. A. Dill and W. C. Swope, Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics J. Chem. Phys., **126** (2007), 155101
- [11] Y. J. Chu and T. H. Liu, On the Shortest Arborescence of a Directed Graph, Science Sinica **14** (1965), 1396 – 1400
- [12] J. P. K. Doye and F. Calvo, Entropic effects on the structure of Lennard-Jones clusters, J. Chem. Phys. **116**, 8307 (2002)

- [13] J. P. K. Doye, M. A. Miller and D. J. Wales, The double-funnel energy landscape of the 38-atom Lennard-Jones cluster, *J. Chem. Phys.* **110** (1999), 6896 – 6906
- [14] W. E and E. Vanden-Eijnden, Towards a theory of transition paths, *J. Stat. Phys.* **123** (2006) 503 – 523
- [15] J. Edmonds, Optimum Branchings, *J. Res. Nat. Bur. Standards* 71B **71B** (1967), 233 – 240
- [16] S. Elmer and S. Park and V. S. Pande, Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular- dynamics simulations in acetonitrile, chloroform, methanol, and water, *J. Chem. Phys.*, **123**, (2005), 114902
- [17] M. I. Freidlin, Sublimiting distributions and stabilization of solutions of parabolic equations with small parameter, *Soviet Math. Dokl.* **18** (1977), 4, 1114 – 1118
- [18] M. I. Freidlin, and A. D. Wentzell, “Random Perturbations of Dynamical Systems”, 3rd Ed, Springer-Verlag Berlin Heidelberg, 2012
- [19] M. I. Freidlin, Quasi-deterministic approximation, metastability and stochastic resonance, *Physica D* **137** (2000), 333 – 352
- [20] K. H. Hoffman and P. Salamon, Bounding the lumping error in Markov chain dynamics, *Appl. Math. Lett.* **22** (2009) 1471 – 1475
- [21] K. H. Hoffman and P. Salamon, Accuracy of coarse grained Markovian dynamics, *Physica A* **390** (2011) 3086-3094
- [22] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, *Proc. Amer. Math. Soc.* **7** (1956), 1, 48 – 50
- [23] J. Kurchan, Six out of equilibrium lectures, Les Houches Summer School, arXiv:0901.1271
- [24] V. A. Mandelshtam and P. A. Frantsuzov, Multiple structural transformations in Lennard-Jones clusters: Generic versus size-specific behavior, *J. Chem. Phys.* **124** (2006), 204511
- [25] V. A. Mandelshtam, P. A. Frantsuzov, and F. Calvo, Structural Transitions and Melting in LJ74-78 Lennard-Jones Clusters from Adaptive Exchange Monte Carlo Simulations, *J. Phys. Chem. A* **110** (2006), 5326 – 5332
- [26] Metzner, P., Schuette, Ch., and Vanden-Eijnden, E.: Transition path theory for Markov jump processes. *SIAM Multiscale Model. Simul.* **7** (2009), 1192 – 1219
- [27] M. Manhart and A. V. Morozov, “Statistical Physics of Evolutionary Trajectories on Fitness Landscapes”, *First-Passage Phenomena and Their Applications*, World Scientific, Singapore, 2014
- [28] M. Manhart and A. V. Morozov, Path-Based Approach to Random Walks on Networks Characterizes How Proteins Evolve New Functions, *PRL* **111** (2013), 088102
- [29] F. Noe, I. Horenko, C. Schuette, and J. C. Smith, Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states, *J. Chem. Phys.* **126**, 975 155102 (2007)
- [30] F. Noe and Ch. Schuette and E. Vanden-Eijnden and L. Reich and T. R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proc. Natl. Acad. Sci. USA*, **106** (2009), 19011
- [31] J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, *Phase changes in 38-atom Lennard-Jones clusters. I. A parallel tempering study in the canonical ensemble*, *J. Chem. Phys.* **112** (2000), 10340
- [32] E. Olivieri and M. E. Vares, “Large Deviations and Metastability”, Cambridge University Press, 2005
- [33] M. Picciani, M. Athenes, J. Kurchan, and J. Tailleur, *Simulating structural transitions by direct transition current sampling: The example of LJ<sub>38</sub>*, *J. Chem. Phys.* **135** (2011), 034108



- [34] J.-H. Prinz and H. Wu and M. Sarich and B. Keller and M. Senne and M. Held and J. D. Chodera and Ch. Schuette and F. Noe, Markov models of molecular kinetics: Generation and validation, *J. Chem. Phys.*, **134** (2011), 174105
- [35] Ch. Schuette, Habilitation thesis, ZIB, Berlin, Germany, 1999, <http://publications.mi.fu-berlin.de/89/1/SC-99-18.pdf>
- [36] Ch. Schuette and F. Noe and J. Lu and M. Sarich and E. Vanden-Eijnden, *J. Chem. Phys.*, **134** (2011), 204105
- [37] P. J. Steinhardt and D. R. Nelson, and M. Ronchetti, Icosahedral Bond Orientational Order in Supercooled Liquids, *Phys. Rev. Lett.*, **47**, 18 (1981) 1297
- [38] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, *Phys. Rev.B*, **28**, 2, (1981) 784 – 805
- [39] W. C. Swope and J. W. Pitera and F. Suits, Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory, *J. Phys. Chem. B*, **108** (2004), 6582
- [40] L. N. Trefethen and D. Bau, Numerical Linear Algebra, SIAM, 1997
- [41] E. Vanden-Eijnden, Transition Path Theory, pp. 91–100 in “An Introduction to Markov State Models and their Application to Long Timescale Molecular Simulation”, Eds. G. R. Bowman, V. S. Pande, F. Noe, Springer, Dordrecht, 2014
- [42] D. J. Wales, Discrete Path Sampling, *Mol. Phys.*, **100** (2002), 3285 – 3306
- [43] D. J. Wales, Energy landscapes: calculating pathways and rates, *International Review in Chemical Physics*, **25**, 1-2 (2006), 237 – 282
- [44] D. J. Wales’s website contains the database for the Lennard-Jones-38 cluster:  
<http://www-wales.ch.cam.ac.uk/examples/PATHSAMPLE/>
- [45] Wales group web site  
<http://www-wales.ch.cam.ac.uk>
- [46] D. J. Wales and J. P. K. Doye, Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters containing up to 110 Atoms, *J. Phys. Chem. A* **101** (1997), 5111 – 5116
- [47] D. J. Wales, M. A. Miller, and T. R. Walsh, *Archetypal energy landscapes*, *Nature* **394** (1998) 758-760
- [48] D. J. Wales, “Energy Landscapes: Applications to Clusters, Biomolecules and Glasses”, Cambridge University Press, 2003
- [49] A. D. Wentzell, *On the asymptotics of eigenvalues of matrices with elements of order  $\exp\{-V_{ij}/(2\epsilon^2)\}$* , (in Russian) *Soviet Math. Dokl.* **202**, No. 2 (1972), 263 – 266